A.F. Bermant

A Course of Mathematical Analysis 2

INTERNATIONAL SERIES OF MONOGRAPHS ON

PURE AND APPLIED MATHEMATICS

GENERAL EDITORS: I. N. SNEDDON, S. ULAM and M. STARK

VOLUME 30

A COURSE OF MATHEMATICAL ANALYSIS

Part II

OTHER TITLES IN THE SERIES

ON PURE AND APPLIED MATHEMATICS

Vol. 1 WALLACE — *An Introduction to Algebraic Topology*

Vol. 2 PEDOE — *Circles*

Vol. 3 SPAIN — *Analytical Conics*

Vol. 4 MIKHLIN — *Integral Equations*

Vol. 5 EGGLESTON — *Problems in Euclidean Space: Application of Convexity*

Vol. 6 WALLACE — *Homology Theory on Algebraic Varieties*

Vol. 7 NOBLE — *Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations*

Vol. 8 MIKUSINSKI — *Operational Calculus*

Vol. 9 HEINE — *Group Theory in Quantum Mechanics*

Vol. 10 BLAND — *The Theory of Linear Viscoelasticity*

Vol. 11 KURTH — *Axiomatics of Classical Statistical Mechanics*

Vol. 12 FUCHS — *Abelian Groups*

Vol. 13 KURATOWSKI — *Introduction to Set Theory and Topology*

Vol. 14 SPAIN — *Analytical Quadrics*

Vol. 15 HARTMAN and MIKUSINSKI — *The Theory of Measure and Lebesgue Integration*

Vol. 16 KULCZYCKI — *Non-Euclidean Geometry*

Vol. 17 KURATOWSKI — *Introduction to Calculus*

Vol. 18 GERONIMUS — *Polynomials Orthogonal on a Circle and Interval*

Vol. 19 ELSGOLC — *Calculus of Variations*

Vol. 20 ALEXITS — *Convergence Problems of Orthogonal Series*

Vol. 21 FUCHS and LEVIN — *Functions of a Complex Variable, Volume II*

Vol. 22 GOODSTEIN — *Fundamental Concepts of Mathematics*

Vol. 23 KEENE — *Abstract Sets and Finite Ordinals*

Vol. 24 DITKIN and PRUDNIKOV — *Operational Calculus in Two Variables and its Applications*

Vol. 25 VEKUA — *Generalized Analytic Functions*

A. F. BERMANT

A COURSE OF MATHEMATICAL ANALYSIS

Part II

Translated by

D. E. BROWN, M. A.

English Translation Edited by

IAN N. SNEDDON

Simson Professor of Mathematics
The University of Glasgow

A Pergamon Press Book

THE MACMILLAN COMPANY
NEW YORK
1963

This book is distributed by
THE MACMILLAN COMPANY · NEW YORK
pursuant to a special agreement with
PERGAMON PRESS LIMITED
Oxford, England

Copyright © 1963
PERGAMON PRESS LTD.

This translation from the Russian has been made from Part II of A. F. Bermant's book entitled "*Kurs matematicheskogo analiza*," published in Moscow 1959 by Gostekhizdat

Library of Congress Card Number 62–9695

MADE IN GREAT BRITAIN

PREFACE

PROFESSOR Bermant's book aims at giving a complete course in mathematical analysis for students of applied science and technology. The first volume covers the requisite work on the theory of functions of one variable. An English translation of this will appear shortly. The present volume is devoted to the theory of functions of several variables, ordinary differential equations, and the elements of the theory of Fourier series. The book contains a wealth of worked examples but there are no problems for solution by the student. This has the advantage that his reading of the subject is not broken up as too often happens in the case of conventional textbooks. To test his comprehension of the subject the student naturally needs to do problems on his own. For this reason a companion book of problems has been prepared by Dr. G. N. Berman. An English translation will appear shortly.

CONTENTS

Preface v

CHAPTER X

Functions of several variables. Differential calculus

1. Functions of several variables 1

136. Concepts. Methods of specifying functions 1
137. Notation for and classification of functions 3
138. The geometrical interpretation of functions 5

2. The elementary investigation of functions 9

139. The domain of definition of a function. The concept of domain 9
140. Limits 14
141. Continuity of a function of several variables. Points of discontinuity 16
142. Some properties of continuous functions. Elementary functions 18
143. The behaviour of a function. Level lines 21

3. Derivatives and differentials of functions of several variables 24

144. Partial derivatives 24
145. Differentials 27
146. Geometrical interpretation of the differential 34
147. Application of the differential to approximations 36
148. Directional derivatives 39
149. Differentiability of functions of two independent variables ... 43

4. Rules for differentiation 46

150. Differentiation of a function of a function 46
151. Implicit functions and their differentiation 51
152. Functions given in the parametric form and their differentiation 55

5. Repeated differentiation 59

153. Derivatives of higher orders 59
154. Differentials of higher orders 63

CHAPTER XI

Applications of the differential calculus

1. Taylor's formula. The extremal of a function of several independent variables 66

155. Taylor's formula and series for functions of several variables... 66
156. Extrema. Necessary conditions 70
157. Problems on absolutely greatest and least values 74
158. Sufficient conditions for an extremum 76
159. Conditional extrema 81

2. Elements of vector analysis 87

160. Vector function of a scalar argument. Differentiation..... 87
161. Gradient 94

3. Curves. Surfaces 97

162. Plane curves 97
163. The envelope of a family of plane curves 99
164. Spatial curves. The helix 105
165. Curvature and torsion. Frenet's trihedral and formulae 112
166. Surfaces 119

CHAPTER XII

Multiple integrals and Iterated integration

1. Double and triple integrals..... 122

167. Problems on volumes. Double integrals..... 122
168. General definition of integral. Triple integrals 126
169. Fundamental properties of double and triple integrals 127
170. Fundamental properties of double and triple integrals (continued). Additive functions of a domain. The Newton–Leibniz formula 130

2. Iterated integration 135

171. Evaluation of double integrals (rectangular domain)..... 135
172. Evaluation of double integrals (arbitrary domain)..... 140
173. Evaluation of triple integrals 148

3. Integrals in polar, cylindrical and spherical co-ordinates 152

174. The double integral in polar co-ordinates 152
175. Triple integrals in cylindrical and spherical co-ordinates 157

4. Applications of double and triple integrals 162

176. Approach for the solution of problems 162
177. Some geometrical problems 166
178. Some problems of statics 168

5. Improper integrals. Integrals dependent on a parameter 172

179. Improper double and triple integrals 172
180. Integrals dependent on a parameter. Leibniz's rule 178

CHAPTER XIII

Line and Surface integrals

1. Line integrals 185

181. Problems concerning work. Integrals over an arc 185
182. Properties, evaluation and applications of line integrals 187

2. Co-ordinate line integrals 191

183. Co-ordinate line integrals 191
184. Component line integrals. Green's formula 197
185. Independence of the integral on the contour of integration ... 202
186. The total differential test. Alternative statements of the fundamental theorem 205
187. Determination of the primitive 210
188. General approach to the solution of problems. Problems of hydrodynamics and thermodynamics 213

3. Surface integrals 219

189. Integrals over a surface area and co-ordinate surface integrals 219
190. Component surface integrals. Stokes' formula 225
191. Ostrogradskii's formula 229

CHAPTER XIV

Differential equations

1. Equations of the first order 233

192. Equations with separable variables 233
193. General concepts 239
194. Equations reducible to equations with separable variables 243
195. Exact differential equations. The integrating factor 249

2. Equations of the first order (continued)..... 254

196. Tangent field. Approximate solutions 254
197. Singular solutions. Clairaut's equation..... 261
198. Orthogonal and isogonal trajectories 266

3. Equations of the second and higher orders 269

199. General concepts 269
200. Particular cases..... 272
201. Approximate solutions 277

4. Linear equations 281

202. Homogeneous equations 281
203. Non-homogeneous equations 289

5. Linear equations with constant coefficients..... 294

204. Homogeneous equations..... 294
205. Non-homogeneous equations 299
206. General formula for the solution of a non-homogeneous equation 304
207. Vibrations. Resonance 308

6. Supplementary problems 313

208. Some linear equations leading to equations with constant coefficients 313
209. Systems of differential equations..... 315

CHAPTER XV

Trigonometric Series

1. Trigonometric polynomials 320

210. The problem 320
211. Fourier coefficients and their properties 323

2. Fourier series 329

212. Fundamental theorems 329
213. Fourier series in an arbitrary interval. Incomplete series 332
214. Examples 335
215. Uniform convergence of Fourier series. Convergence "in mean" 342
216. The Parseval-Lyapunov theorem 347

3. Krylov's method. Practical harmonic analysis 350

217. Order of the coefficients 350
218. Krylov's method of improving the convergence of trigonometric series 353
219. Examples 355
220. Practical harmonic analysis. Templates 360

Index 367

CHAPTER X

FUNCTIONS OF SEVERAL VARIABLES DIFFERENTIAL CALCULUS

1. Functions of Several Variables

136. Concepts. Methods of specifying functions

Definition. **A quantity such that one or several definite values of it correspond to every given pair of values of two variables is termed a *function of the two variables.***

The two variables which vary arbitrarily are termed independent variables.

For instance, the area S of a rectangle is a function of two variables which can vary independently, namely the sides a and b. The expression for this function is

$$S = ab.$$

The volume v of an ideal gas is a function of the pressure p and the temperature T. The expression for v in terms of p and T is found from the relationship between v, p and T; this is the familiar Mendeleev–Clapeyron equation

$$v = \frac{RT}{p} \quad (R = \text{const.}).$$

To specify a function of two variables means to indicate the system of pairs of values which the independent variables take and the method by which the corresponding values of the function are found for the given values of the variables. Just as in the case of a function of a single variable, the most important methods of specifying a function of two variables are the *tabular*, *analytic* and *graphical* methods.

The tabular method gives the corresponding values of the function for a certain number of pairs of values of the independent

variables. This is usually done by means of a table with double entries, which must be arranged so as to be easily read: as e.g. by the following table, establishing the relationship between the efficiency of a helical gear η and the friction coefficient μ and angle of lift α of the helix (BERLOV, *Machine Details* (*Detali Mashin*)):

η \ α	$\frac{\pi}{36}$	$\frac{\pi}{18}$	$\frac{\pi}{12}$	$\frac{\pi}{9}$	$\frac{5\pi}{36}$
0·01	0·897	0·945	0·961	0·970	0·974
0·02	0·812	0·895	0·925	0·941	0·950
0·03	0·743	0·850	0·892	0·914	0·927
0·04	0·683	0·809	0·861	0·888	0·904
0·05	0·633	0·772	0·831	0·863	0·882

As in the case of functions of one variable, the most important method of all of specifying a function of two variables is the analytic method, i.e. with the aid of a formula.

The analytic method indicates the sequence of mathematical operations, with the aid of which the values of the independent variables are connected with the corresponding values of the function, i.e. in other words, a formula is given in which three variables play a part (see Sec. 10).

For instance, each of the expressions

$$z = ax + by + c, \quad z = \frac{xy}{x^2 + y^2}, \quad z = \frac{\sin(2x + 3y)}{\sqrt{1 + (x - y)^2}}$$

gives z as a function of x and y.

The entire equipment of mathematical analysis is in fact adapted to the analytic specification of functions.

We shall touch on the graphical specification of functions of two variables in Sec. 138.

A function of any number of independent variables may be defined in precisely the same way as for a function of two variables.

Definition. **A quantity is termed a *function of n independent variables* if there correspond to any given set of values of the variables one or several definite values of the quantity.**

For instance, the volume V of a rectangular parallelepiped is a function of three independent variables: the sides a, b, c of the parallelepiped:

$$V = abc.$$

The quantity of heat Q produced by an electric current depends on the voltage E, the current I and time t, the functional relationship between these quantities being given by

$$Q = 0{\cdot}24IEt.$$

A function of n variables ($n > 2$) can also be specified by tables, but this method is extremely laborious even with $n = 3$.

The analytic method of specifying a function of n variables consists in indicating the sequence of mathematical operations which connects the value of the function with the corresponding values of the n independent variables.

For instance, each of the equations

$$u = ax + by + cz + d, \quad u = \frac{xyz}{x^2 + y^2 + z^2}$$

gives u as a function of x, y, z.

A constant can be regarded when necessary *as a function of any number of variables*, i.e. as a function that retains the same value for any set of values of the independent variables.

Almost all our arguments will be carried out in detail only for functions of two independent variables, bearing in mind that, as a rule, the corresponding arguments for a function of more than two variables are precisely similar.

137. Notation for and Classification of Functions. To express the fact that u is a function of variables x, y, z, t, ..., we write for instance

$$u = f(x, y, z, t, \ldots);$$

all the variables upon which the function depends are put in brackets after the symbol for the function (for which any letter may be used, apart from f).

If we keep the value of one of the variables, say x, constant, $x = a$, and continue to regard the other, y, as variable, the function $z = f(x, y)$ becomes the function $z = f(a, y)$ of the single variable y. Similarly, if y retains the constant value $y = b$, the function $z = f(x, b)$ of the single variable x is obtained.

A function of two independent variables is said to be single-valued if a *single* value of the function corresponds to any given pair of values of the variables; whereas if more values than one of the function correspond to certain pairs of the variables, the function is said to be many-valued.

For example, the functions

$$z = ax + by + c, \quad z = e^{xy} \sin(x + y)$$

are single-valued; whereas

$$z = \pm \sqrt{x^2 + y^2}, \quad z = \operatorname{Arcsin}(x - y)$$

are many-valued functions, the first being two-valued and the second "infinitely valued".

Only single-valued functions are considered below unless there is some proviso to the contrary. Whenever mathematical analysis is applied to some "many-valued expression" a special agreement must be reached in each case regarding the choice of the single-valued branch of the function concerned.

Functions of several variables can generally be regarded as functions of a function of the independent variables. Suppose that u is expressed as a function of arguments $t, v, w, \ldots$:

$$u = \varphi(t, v, w, \ldots),$$

whilst $t, v, w, \ldots$ are in turn functions of independent variables $x, y, z, \ldots$:

$$t = \psi(x, y, z, \ldots), \quad v = \xi(x, y, z, \ldots), \quad w = \eta(x, y, z, \ldots), \ldots$$

Then u is a function of the independent variables $x, y, z, \ldots$, specified by means of a chain of functions. The function

$$\begin{aligned} u &= \varphi[\psi(x, y, z, \ldots), \xi(x, y, z, \ldots), \eta(x, y, z, \ldots), \ldots] \\ &= F(x, y, z, \ldots) \end{aligned}$$

is termed a function of a function of independent variables x, y, z, ...

For example, $z = e^t \sin v$, where $t = xy$, $v = x + y$, is a function of a function of the two independent variables x and y.

A function of a function u of independent variables $x, y, \ldots$ can be specified by means of two or more intermediate links.

A function of a function, expressed directly in terms of a certain number of intermediate arguments, may be a function of a different number of *independent* variables. If, for instance, arguments u and v of the function

$$z = f(u, v)$$

are functions of the independent variable x, i.e. $u = \varphi(x)$, $v = \psi(x)$, then z is a function of a function of a *single* independent variable:

$$z = f[\varphi(x), \psi(x)] = F(x).$$

The functions of a function most commonly encountered in analysis and its applications are those in which the links are basic elementary functions of the independent variables.

A function of several variables is said to be rational if its value can be obtained from the values of the independent variables with the aid of the operations of addition, subtraction, multiplication, division and raising to an integral power. For example,

$$z = (3x^2 - 5xy + 1)/(x^2y - y^2 - 1)$$

is a rational function of x and y.

A rational function is described as integral or as a polynomial if division by an expression containing the independent variables is not required when working out its values. For example,

$$z = 3x^3y - 5x^2y^2 - y^3 + xy - 1$$

is a polynomial.

The first-degree polynomial or linear function has a special significance due to its simplicity:

$$z = ax + by + c, \quad u = ax + by + cz + d \quad \text{etc.}$$

($a, b, c, \ldots$ are constants; $x, y, \ldots$ are independent variables).

138. The Geometrical Interpretation of Functions. We made use of a *plane* for the geometrical interpretation of functions of a single independent variable. We shall make use of *space* for the geometrical interpretation of functions of two independent variables. The equations defining such functions connect three variables, and we can represent these, say, with the aid of three axes of a system of rectangular Cartesian co-ordinates in space (Fig. 1). It must be noted that there are two essentially different systems of rectangular Cartesian co-ordinates in space (as in a plane). These systems differ in the *orientation* of the co-ordinate axes. Figure 1 *a* illustrates the so-called right-handed system, and Fig. 1 *b* the left-handed system.

If we view the Oxy plane from the side of the positive semi-axis Oz, the direction of rotation (via the shortest path) from the positive semi-axis Ox to the positive semi-axis Oy is anticlockwise in the right-handed system. The direction of rotation is clockwise in the left-handed system. Three fingers in the order: thumb, index, middle finger, corresponding to the sequence of axes: Ox, Oy, Oz, form a right-handed system on the right hand, and a left-handed system on the left hand.

In most problems of analysis it is of no consequence which system of co-ordinates (right- or left-handed) is employed. However, there are certain problems in which there is a preferred orientation in space, due to the system of co-ordinates chosen. Since a right-handed system is normally used in a plane (viewing the plane from above), we shall also use a right-handed system in space (Fig. 1 *a*).

Given a function z of two independent variables x and y, its geometrical shape, or what we can call its graph, is the *locus of points whose first two co-ordinates are the values of the independent*

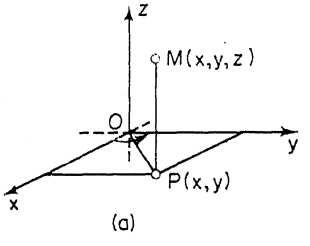

FIG. 1 a

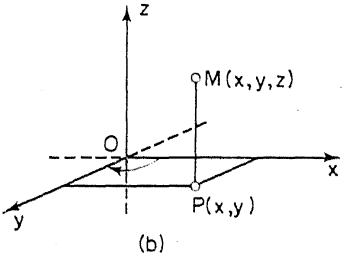

FIG. 1 b

variables whilst the third co-ordinate is the corresponding value of the function. Every pair of values of the independent variables x and y defines a point $P(x, y)$ in the Oxy plane; the end of the perpendicular to the Oxy plane at the point P which expresses the value of function z is a point $M(x, y, z)$, representing the triad of the corresponding values of the independent variables and function (Fig. 1). The set of all such points M in space, corresponding to all possible pairs of values of x and y, in other words, to all possible positions of the point P in the Oxy plane, in fact forms the geometrical shape (graph) of the function. In the cases commonly encountered this shape is a surface.

Conversely, specifying a surface in space and a co-ordinate system establishes z as a function of x and y. For, the z co-ordinate of a point of this surface is defined by the abscissa and ordinate, i.e. is a function of the abscissa and ordinate. We say in this case that the function $z = f(x, y)$ is *specified graphically.*

If the function is specified by the analytic expression $z = f(x, y)$, this formula is the equation of the surface formed by the graph of the given function.

For example, the graph of the function

$$z = \frac{c}{ab}\sqrt{a^2b^2 - b^2x^2 - a^2y^2}$$

is the upper half of the ellipsoid (Fig. 2) with centre at the origin and semi-axes a, b, c, lying along the Ox, Oy, Oz axes. In fact, the formula

$$z = \pm \frac{c}{ab} \sqrt{a^2b^2 - b^2x^2 - a^2y^2} \quad \left(\text{or} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\right)$$

is the equation of this ellipsoid. The graph of the function

$$z = x^2 + y^2$$

is the paraboloid of revolution (Fig. 3), whilst the graph of the linear function

$$z = ax + by + c$$

is a plane; in particular, the plane corresponding to the equation $z = c$ (c = const.) is parallel to the Oxy plane.

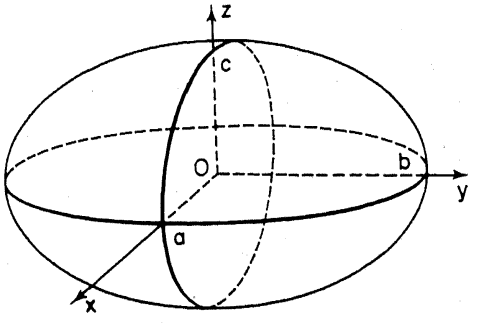

FIG. 2

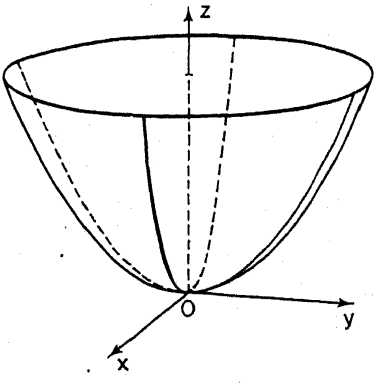

FIG. 3

There is thus a one-to-one correspondence between functions of two variables and surfaces in space, just as there is between functions of a single variable and curves on a plane: *every function is usually represented by a certain surface, and every surface is defined by a certain function.*

The fact that the function is single-valued is expressed geometrically by the fact that any straight line perpendicular to the Oxy plane intersects the surface representing the function in not more than *one* point.

We can now offer a unified geometrical interpretation for a function of any number of arguments. It is assumed here, as previously mentioned, that the functions are single-valued.

We take the function $y = f(x)$. Its argument x is represented by a point $P(x)$ moving along a *straight line* — the Ox axis, whilst

the corresponding value of the function y measures the length of the perpendicular to the Ox axis at the point P.

In the case of the function $z = f(x, y)$ the set of the two arguments x and y is also represented by a *point* $P(x, y)$, which, however, now moves in a *plane* — the Oxy plane, whilst the corresponding value of the function z measures the length of the perpendicular to the Oxy plane at the point P.

Definition. **If there is a definite value of a certain magnitude corresponding to any given position of a point P, this magnitude is said to be a *function of the variable point* P.**

Hence the function in both the above cases can be regarded as a function of the point P:

$$y = f(P), \quad z = f(P),$$

the difference being that point P — the argument of the function — moves along a straight line in the first case and over a plane in the second.

A function of three independent variables can also be regarded as a function of a point, i.e. a point $P(x, y, z)$ moving in space of three dimensions (ordinary space):

$$u = f(P) = f(x, y, z).$$

As regards the geometrical representation of the function itself, a method similar to that for functions of one and two variables would require a space of four dimensions. When speaking of a space of four dimensions, we imply the system of all possible sets of four values x, y, z and u, each set being a point of the space. For every given position of the point $P(x, y, z)$ (the argument of the function) in three-dimensional space there now corresponds the point $M(x, y, z, u)$ in four-dimensional space, representing the set of the values of the arguments and the corresponding value of the function.

The treatment is similar for functions of four, five and in general any number of independent variables.

A function u of n independent variables $x, y, z, \ldots, t$ is a function of a point P moving in space of n dimensions:

$$u = f(P) = f(x, y, z, \ldots, t).$$

It is not easy to make direct use of the geometrical interpretation even in the case of functions of two independent variables, due to the limitations of our spatial intuition and the impossibility of drawing in space. Usually, however, the investigation of a function of two or more variables reduces to the investigation of a function of a single variable (see Sec. 143).

Let one of the independent variables, say x, of a given function $z = f(x, y)$ remain constant: $x = x_0$. This limits the freedom of movement of the point $P(x, y)$ —it can only move along the straight line $x = x_0$ in the Oxy plane. We obtain the function of one variable:

$$z = f(P) = f(x_0, y);$$

the condition $x = x_0$ implies geometrically that the surface representing the function $z = f(x, y)$ is intersected by the plane

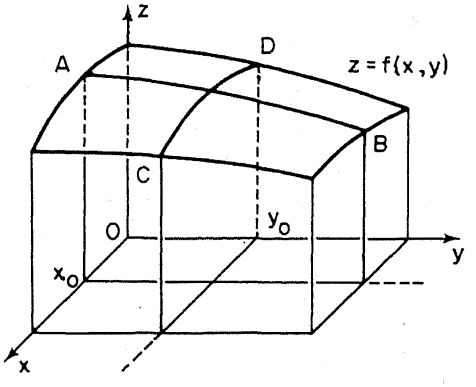

FIG. 4

$x = x_0$; let the plane curve AB be obtained by this means (Fig. 4). The equation $z = f(x_0, y)$ is the equation of the orthogonal projection of the curve AB on the plane Oyz.

Similarly the argument P of the function

$$z = f(P) = f(x, y_0)$$

can be displaced only along the straight line $y = y_0$ in the Oxy plane. The graph of this function of one variable is the orthogonal projection on to the Oxz plane of the curve CD (Fig. 4) —the intersection of the surface with the plane $y = y_0$.

2. The Elementary Investigation of Functions

139. The Domain of Definition of a Function. The Concept of Domain. I. We say that a function $z = f(x, y)$ is defined at a point $P(x, y)$ of the Oxy plane if the value (or values) of function z is defined at this point in accordance with some given rule.

Definition. **The *domain of definition* (or *of existence*) of a function $z = f(x, y)$ is the set of all points of the Oxy plane at which the function is defined.**

The domain of definition may be any set of points of the plane, but the functions generally considered in analysis are defined in a certain part of the plane bounded by a curve or curves (excluding possibly individual points of this part of the plane, or individual curves, where the function is not defined).

Definition. **A part of the plane bounded by one or more curves** (which may stretch to infinity) **is termed a *domain*, the curves that bound it being its *boundary*.**

If the boundary is included with the domain, the latter is said to be closed*, whilst if the boundary is excluded the domain is

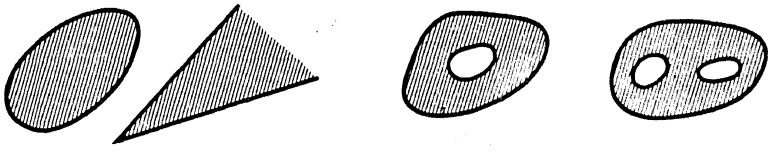

FIG. 5

said to be open. (Where it is unnecessary to distinguish between these cases we shall simply speak of a "domain.")

A domain is described as connected if *any* two points of it can be joined by a continuous curve (e.g. a step line) lying wholly within the domain.

A connected domain is described as simply connected (Fig. 5*a*) if its boundary consists of a single continuous curve. A simply-connected domain is also characterized by the fact that any closed curve lying in the domain can be contracted to a point by continuous deformation, without at any time passing through a point not belonging to the domain. The circle represents a typical example of a simply-connected domain. If the boundary of a connected domain consists of two separate continuous curves (in particular, points), the domain is said to be doubly connected (Fig. 5*b*). A typical example of a doubly connected domain is provided by the annular ring (in particular, by the circle with its centre removed).

A connected domain is said to be triply connected if its boundary consists of three separate continuous curves (Fig. 5*c*), and so on.

Definition. **An *r-neighbourhood* of a point is the circle of radius *r* with centre at the point.**

* The boundary of a closed domain cannot extend to infinity.

As in the case of intervals on the Ox axis, domains in the Oxy plane can be specified with the aid of one or more inequalities.

Example 1. We take the rectangle whose sides lie on the straight lines $x = a$, $x = b$, $y = c$, $y = d$ (Fig. 6). The domain D bounded by this rectangle can be specified by means of the inequalities

$$a \leqslant x \leqslant b, \quad c \leqslant y \leqslant d \qquad \text{(closed)}$$

or

$$a < x < b, \quad c < y < d \qquad \text{(open)}.$$

In fact, the co-ordinates of any point belonging to domain D satisfy these inequalities, and conversely, any two numbers x and

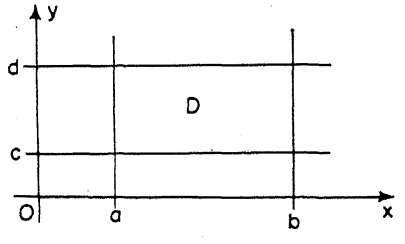

FIG. 6

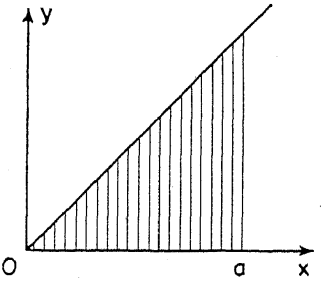

FIG. 7

y satisfying these inequalities are co-ordinates of some point of domain D.

Example 2. Let the domain D be bounded by a circle with centre at the origin and radius r. Domain D is specified by the inequality

$$x^2 + y^2 \leqslant r^2 \qquad \text{(closed)}$$

or

$$x^2 + y^2 < r^2 \qquad \text{(open)}.$$

The part of the circle lying in the first quadrant is specified by the system of inequalities

$$x^2 + y^2 \leqslant r^2, \quad x \geqslant 0, \quad y \geqslant 0 \quad \text{or} \quad x^2 + y^2 < r^2, \quad x > 0, \quad y > 0.$$

The annular ring (doubly-connected domain) bounded by circles with common centre at the origin and radii r and R ($r < R$) can be specified with the aid of the inequalities $r^2 \leqq x^2 + y^2 \leqq R^2$ or $r^2 < x^2 + y^2 < R^2$.

Example 3. The domain D (Fig. 7) included between the Ox axis, the bisector of the first quadrant and the straight line $x = a$ (right-angled triangle) is specified by the inequalities

$$0 \leqslant y \leqslant x, \quad 0 \leqslant x \leqslant a \quad \text{or} \quad 0 < y < x, \quad 0 < x < a.$$

Example 4. We take as domain D the whole of the half-plane lying to the right of the Oy axis (Fig. 8). It is specified by the inequality $x \geqq 0$ (or $x > 0$). The inequalities $x \geqq 0$, $y \geqq 0$ (or $x > 0$, $y > 0$) specify the first quadrant of the Oxy plane. (The whole of the Oxy plane is specified by the inequalities $-\infty < x < \infty$, $-\infty < y < \infty$.)

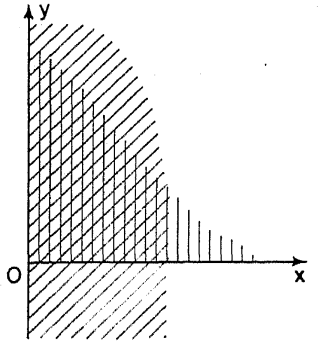

FIG. 8

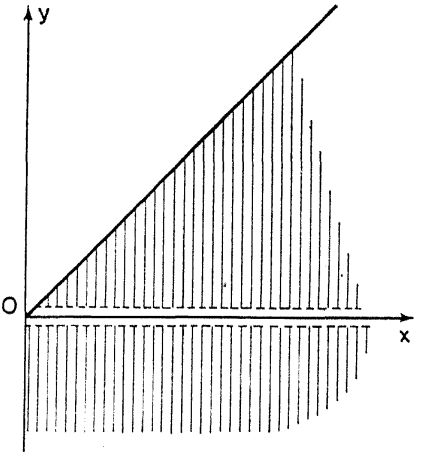

FIG. 9

A part of space bounded by one or more surfaces is termed a domain of space. As with domains of a plane, domains of space can be specified with the aid of inequalities.

The concepts introduced above for functions of two independent variables can in general be carried over without significant changes to functions of three independent variables.

The domain of definition of a function of three variables is usually some domain of space, excluding possibly individual points, curves and surfaces of the domain where the function is not defined.

II. *Definition.* **The *domain of definiteness* of an analytic expression is the set of all points at which the expression has a definite numerical (real) value.**

For instance, the domain of definiteness of the expression

$$\sqrt{r^2 - x^2 - y^2}$$

is the circle

$$x^2 + y^2 \leqslant r^2;$$

the domain of definiteness of any polynomial is the whole of the Oxy plane; the domain of definiteness of the expression

$$\sqrt{x - y}\ \ln(xy^2)$$

is the domain (shaded in Fig. 9) specified by the inequalities

$$y \leqslant x, \quad x > 0, \quad y \neq 0.$$

If a function $z = f(x, y)$ is specified by a single analytic expression without any subsidiary conditions, we take as the domain of definition of the function the "domain of definiteness" of the analytic expression.

Thus the domain of definition of the function

$$z = \frac{\sqrt{1 - x^2 - y^2}}{x^2 + y^2}$$

is the circle $x^2 + y^2 \leqq 1$, excluding the point $(0, 0)$; the function

$$z = \frac{\sin(xy)}{x - y}$$

is defined throughout the Oxy plane, except for the straight line $y = x$.

If $f(x, y)$ is given by an analytic expression and certain *auxiliary conditions*, either assigned earlier or following from the nature of the problem, the domain of definition of $f(x, y)$ will in general differ from the domain of definiteness of the analytic expression. For instance, let $f(x, y)$ be given as follows: it is equal to $(x - y) \sin [1/(x - y)]$ for all x and y not equal to each other, and is equal to zero for equal values of x and y:

$$f(x, y) = \begin{cases} (x - y) \sin \dfrac{1}{x - y} & (x \neq y), \\ 0 & (x = y). \end{cases}$$

The domain of definiteness of the expression $(x - y) \sin [1/(x - y)]$ is the Oxy plane excluding the straight line $y = x$ (the argument of the sine becomes meaningless at such points), whilst the domain of definition of the function $f(x, y)$ is the whole of the Oxy plane.

On the other hand, concrete problems sometimes lead to functions written as analytic expressions which are defined in a wider domain than that permitted by the conditions of the problem. We can take as an example the function $f(x, y)$ expressing the area of a rectangle as a function of its sides x and y:

$$f(x, y) = xy.$$

The expression xy is defined throughout the Oxy plane, but it would be beyond the meaning of the present problem to consider negative values of x and y. The domain of definition of the present function can therefore only be the first quadrant: $x \geqq 0$, $y \geqq 0$.

140. Limits. If we regard a function $z = f(x, y)$ of two independent variables x and y as a function of a point $P(x, y)$ on the plane Oxy, the following definition can be given of the limit of the function, analogous to the definition of limit for a function of one variable.

Definition 1. **The number A is the *limit of the function* $z = f(P) = f(x, y)$ as $P(x, y) \to P_0(x_0, y_0)$ if, given any positive number ε, a corresponding positive number δ can be found such that, for all points $P(x, y)$ differing from $P_0(x_0, y_0)$ for which $\varrho = PP_0 < \delta$, we have**

$$|A - f(P)| < \varepsilon.$$

We write this as:

$$\lim_{P \to P_0} f(P) = A. \qquad (*)$$

This definition may be briefly expressed as:

The number A is termed the limit of the function $z = f(P)$ *as* $P \to P_0$ *if the difference* $A - f(P)$ *is infinitesimal when the distance* $\varrho = PP_0$ *is infinitesimal.*

Having clearly indicated the variables on which the function depends, we can also say that:

The number A is the limit of the function $z = f(x, y)$ *as* $x \to x_0$ *and* $y \to y_0$:

$$\lim_{\substack{x \to y_0 \\ y \to y_0}} f(x, y) = A,$$

if, given any positive number ε, *a corresponding positive number* δ *can be found such that, for all* x *and* y *for which* $|x - x_0| < \delta$ *and* $|y - y_0| < \delta$, *where either* $x \neq x_0$ *or* $y \neq y_0$, *we have*

$$|A - f(x, y)| < \varepsilon.$$

The expression (*) is to be understood in the sense that $f(P)$ tends to A *independently of how* $P(x, y)$ approaches $P_0(x_0, y_0)$. If it is intended that $P(x, y)$ should tend to $P_0(x_0, y_0)$ in a perfectly definite manner, this must be mentioned. For instance, we may be interested in the limit to which $f(P)$ tends when the point $P(x, y)$ approaches the point $P_0(x_0, y_0)$ only along straight lines parallel to the axes or in general along any given curve.

As in the case of a single independent variable, the function $f(P)$ may not itself be defined at the limit point P_0; this does not exclude the possibility that $f(P)$ tends to a limit when $P \to P_0$. For example, the function

$$z = \frac{\sin(x - y)}{x - y}$$

is not defined at the point $P_0(2, 2)$. But $z \to 1$ as the point $P(x, y)$ tends to the point $P_0(2, 2)$ along any path not containing points of the straight line $y = x$.

We now give a definition of the limit of the function of two variables $z = f(P)$ in the case when the point $P(x, y)$ tends to infinity, and the definition of the unbounded growth of the function $f(P)$ in various cases.

Definition 2. *The number A is termed the limit of $f(P)$ as $P \to \infty$:*

$$\lim_{P \to \infty} f(P) = A,$$

if, given any number $\varepsilon > 0$, a number $N > 0$ can be found such that, for all points $P(x, y)$ for which $\varrho = PO > N$, we have

$$|A - f(P)| < \varepsilon.$$

Definition 3. *The function $f(P)$ increases indefinitely as $P \to P_0$:*

$$\lim_{P \to P_0} f(P) = \infty,$$

if, given any positive number M, a $\delta > 0$ can be found such that, for all points $P(x, y)$ differing from P_0 for which $\varrho = PP_0 < \delta$, we have

$$|f(P)| > M.$$

Definition 4. *A function $f(P)$ increases indefinitely as $P \to \infty$:*

$$\lim_{P \to \infty} f(P) = \infty,$$

if, given any positive M, an $N > 0$ can be found such that, for all points $P(x, y)$ for which $\varrho = PO > N$, we have

$$|f(P)| > M.$$

As earlier, if the movement of the point $P(x, y)$ is restricted in some way or if $f(P)$ tends to an infinity with a definite sign, this must be mentioned.

To define a limit for the function

$$u = f(P) = f(x, y, z, \ldots, t)$$

of n independent variables we require a word for word repetition of the definitions for the case $n = 2$, except for replacing the expression for the distance PP_0 between points P and P_0 on a plane

by the expression for the distance PP_0 between points $P(x, y, z \ldots, t)$ and $P_0(x_0, y_0, z_0, \ldots, t_0)$ in space of n dimensions. This distance is given by

$$\varrho = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 + \cdots + (t - t_0)^2}.$$

141. Continuity of a Function of Several Variables. Points of Discontinuity. Let the point $P_0(x_0, y_0)$ belong to the domain of definition of the function $f(P)$.

Definition. **A function $z = f(P) = f(x, y)$ is said to be continuous at the point P_0 (x_0, y_0) (or at $x = x_0$, $y = y_0$) if**

$$\lim_{P \to P_0} f(P) = f(P_0),$$

where $P(x, y)$ can tend to P_0 (x_0, y_0) in any manner.

We can express this alternatively as:

A function $z = f(P)$ is said to be continuous at a point P_0 *if, for an infinitesimal displacement of the point P $(\Delta\varrho = PP_0 \to 0)$, there is a corresponding infinitesimal variation of the function $(\Delta z = f(P) - f(P_0) \to 0)$.*

Three requirements have to be satisfied for continuity of a function $z = f(P)$ at a point P_0:

(1) $f(P)$ must be defined in some neighbourhood of the point P_0 (or in a closed domain with P_0 on its boundary);

(2) $f(P)$ must have a limit when P tends to P_0 in an arbitrary manner (if P_0 is on the boundary of the domain, P can approach P_0 only from within the boundary);

(3) this limit must coincide with the value of the function at the point P_0.

Definition. **A function which is continuous at every point of a domain is said to be *continuous in the domain*.**

The continuity of a function $z = f(x, y)$ implies geometrically that the z co-ordinates of its graph corresponding to two points of the Oxy plane differ from each other by as little as may be desired if the distance between these two points is sufficiently small. Hence the graph of a continuous function consists of a continuous surface with no breaks.

The continuity of a function of any number of independent variables is similarly defined.

Definition. **P_0 is said to be a *point of discontinuity* of a function $z = f(P)$ if $f(P)$ is defined in some neighbourhood of this point with the exception of the point P_0 itself or of some curve passing through P_0,**

or if $f(P)$ is defined at every point of some neighbourhood of the point P_0 but does not satisfy the second or third of the above requirements*.

The points of discontinuity of a function $z = f(x, y)$ may form a curve. Such a curve is termed a curve of discontinuity of the function. The function $f(P)$ is said to be discontinuous at a point of discontinuity.

Some examples will be mentioned of discontinuous functions and points of discontinuity:

(1) The function $z = \sin 1/\sqrt{x^2 + y^2}$ is defined throughout the Oxy plane except for the point $P_0(0, 0)$; the function is discontinuous at this point. It is continuous at all other points of the plane. The function is represented geometrically by the surface obtained by rotating the graph of the function $z = \sin 1/x$, $x > 0$ about the Oz axis.

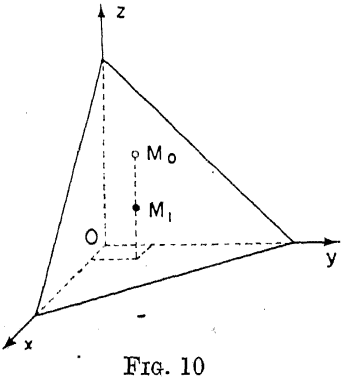

FIG. 10

The function $z = \sin 1/(\sqrt{x^2 + y^2} - 1)$ is discontinuous at every point of the circle $x^2 + y^2 = 1$. This circle is a curve of discontinuity of the function.

(2) We define a function $z = f(P)$ as follows: $f(P)$ is equal to $3 - x - y$ at every point $P(x, y)$ of the Oxy plane except for the point $P_0(1, 1)$ where its value is $\frac{1}{2}$. This function is discontinuous at the point $P_0(1, 1)$, since the third requirement is not satisfied: no matter how $P(x, y)$ tends to $P_0(1, 1)$, the function tends to 1 and not to $\frac{1}{2}$, as would be required for continuity. The graph of this function (Fig. 10) consists of the whole of the plane $z = 3 - x - y$ without the point $M_0(1, 1, 1)$, instead of which we have the point $M_1(1, 1, \frac{1}{2})$ belonging to the graph.

Points of discontinuity of a function of n independent variables are defined as in the case $n = 2$.

It may be remarked that a function $z = f(x, y)$ can be continuous at a point with respect to each of the independent variables separately and yet be discontinuous at the point with respect to their aggregate, i.e. as a function of an *arbitrarily moving* point $P(x, y)$.

We can take as an illustration:

$$f(x, y) = \frac{2xy}{x^2 + y^2} \quad \text{for} \quad x \neq 0, \quad y \neq 0,$$
$$f(0, 0) = 0.$$

* The limits of sets of points of discontinuity of a function are also regarded as points of discontinuity of the function.

This function is defined throughout the plane and is discontinuous at the origin. In fact, at every point of the straight line $y = kx$, where k is any number, $f(x, y) = 2kx^2/x^2(1 + k^2) = 2k/(1 + k^2)$, which means that the function has a limit equal to the number $2k/(1 + k^2)$ when the argument — the point $P(x, y)$ — tends to $P_0(0, 0)$ along the straight line $y = kx$. Thus the limit of $f(x, y)$ as x and y tend to zero simultaneously can be any number lying between -1 and $+1$ (since the equation $2k/(1 + k^2) = \mu$, where $-1 < \mu < 1$, always has two real roots); this number depends on the path along which the point $P(x, y)$ approaches the point $P_0(0, 0)$, which proves that the function is discontinuous at $P_0(0, 0)$.

At the same time the given function, regarded as a function of one of its arguments (i.e. with a constant value of the other), is continuous throughout any straight line along which the point P moves. For instance, let $y = y_0$; we now obtain a function of x which we shall write as $\varphi_1(x)$:

$$\varphi_1(x) = f(x, y_0) = \frac{2xy_0}{x^2 + y_0^2}.$$

If $y_0 \neq 0$, the continuity of this function is obvious. Whilst if $y_0 = 0$, $\varphi_1(x) \equiv 0$ for $x \neq 0$, and with $x = 0$ the function is equal to zero by virtue of the auxiliary condition that $f(0, 0) = 0$. Being identically zero, the function is continuous at every point of the Ox axis, including $P_0(0, 0)$. Similarly, the function

$$\varphi_2(y) = f(x_0, y) = \frac{2x_0y}{x_0^2 + y^2}$$

is continuous for any y.

On the other hand, if a function $z = f(x, y)$ is continuous with respect to its system of independent variables, it is continuous also with respect to each of the independent variables separately.

142. Some Properties of Continuous Functions. Elementary Functions

I. Let the function $f(P) = f(x, y)$ be continuous in a closed domain D. It now possesses the properties indicated by the following theorems (cf. Sec. 37).

THEOREM 1. *The function $f(P)$ attains a maximum and minimum value at least once in the domain D.*

Theorem 1 asserts the existence in domain D of at least one point $P_1(\xi_1, \eta_1)$ and one point $P_2(\xi_2, \eta_2)$ such that the values of the function at these points are respectively its greatest and least values throughout the domain:

$$f(P_2) \leqslant f(P) \leqslant f(P_1), \quad (f(\xi_2, \eta_2) \leqslant f(x\ y) \leqslant f(\xi_1\ \eta_1))$$

for all points $P(x, y)$ of the domain.

THEOREM 2. *The function $f(P)$ attains at least once in domain D any value intermediate to two values taken by the function in D.*

Let the function take the values

$$f(P_1) = f(x_1, y_1) = M_1, \quad f(P_2) = f(x_2, y_2) = M_2$$

at points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ in D and let say $M_1 < M_2$. We take any number M intermediate to M_1 and M_2, $M_1 < M < M_2$. Theorem 2 says that at least one point $P(\xi, \eta)$ exists in D such that

$$f(P) = f(\xi, \eta) = M.$$

In particular, if $f(x, y)$ takes both negative and positive values in domain D, the function vanishes at at least one point $P(\xi, \eta)$ of the domain:

$$f(P) = f(\xi, \eta) = 0.$$

THEOREM 3. *Function* $f(P)$ is *uniformly continuous in domain* D, i.e. *given any positive* ε, *a positive* δ *can be found such that*

$$|f(P) - f(P_0)| < \varepsilon,$$

provided $PP_0 < \delta$, *independently of the position of the point* P_0 *in domain* D.

The proofs of these theorems will be found in more complete works on analysis*.

Theorems 1–3 can be carried over word for word to functions of any number of independent variables.

II. The rules for passage to the limit in a sum, product and quotient of functions also hold in the case of functions of n independent variables. The following theorems are easily proved with the aid of these rules, as in the case of a single independent variable.

THEOREM 1. *The sum and product of a finite number of functions which are continuous at a given point are also continuous at this point.*

THEOREM 2. *The quotient of two functions continuous at a given point is continuous at this point provided the denominator does not vanish.*

THEOREM 3. *A function of a function is continuous if it is composed of a finite number of continuous functions.*

Proof. Suppose we have the function of a function

$$u = \varphi(t, v, w, \ldots),$$

where $t, v, w, \ldots$ are functions of the independent variables $x, y, z, \ldots$

* G.M. FIKHTENGOLTS, *Course of Differential and Integral Calculus (Kurs differentsialnogo i integralnogo ischisleniya)*, Vols. I–III, Gost., 1947; R. COURANT, *Course of Differential and Integral Calculus*, Parts 1 and 2, 1931.

Let φ, ψ, ξ, η, ... all be continuous functions of their arguments. Then

$$u = \varphi(t, v, w, \ldots) = F(x, y, z, \ldots)$$

is a continuous function of the point $P(x, y, z, \ldots)$.

For, in view of the continuity of functions ψ, ξ, η, ..., an infinitesimal displacement of the point $P(x, y, z, \ldots)$ produces infinitesimal variations of variables t, v, w, ..., which in turn produce an infinitesimal increment in the variable u by virtue of the continuity of function φ. As a result, an infinitesimal variation in u corresponds to an infinitesimal displacement of the point $P(x, y, z, \ldots)$, which characterizes the continuity of u as a function of the point $P(x, y, z, \ldots)$.

III. The functions of several variables generally encountered in a general course of mathematical analysis and in the applied sciences are *elementary. As in the case of a single independent variable, an* elementary function *is one which can be specified by means of a single analytic expression made up from the independent variables and constants with the aid of a finite number of basic elementary functions and arithmetic operations.*

An elementary function can only have discontinuities at points where one or more of the elements composing it (or the function itself) are indeterminate or where the denominators of fractions in the expression for the function vanish.

Hence *the domain of continuity of an elementary function (the aggregate of all points at which the function is continuous) coincides strictly with its domain of definition.*

This leads to a rule which is simple and convenient to use.

RULE FOR PASSAGE TO THE LIMIT. *To find the limit of an elementary function we only need to evaluate the function at the limit point.*

In other words, *we have to replace the independent variables in the expression for the function by the limits to which these tend.*

For we have, since the function is continuous,

$$\lim_{P \to P_0} f(x, y, z, \ldots, t) = f(x_0, y_0, z_0, \ldots, t_0),$$

i.e. $$\lim f(x, y, z, \ldots, t) = f(\lim x, \lim y, \lim z, \ldots, \lim t),$$

which can be written as

$$\lim f(P) = f(\lim P).$$

Thus *the symbol for the limit and the symbol for a continuous function can be interchanged.*

143. The Behaviour of a Function. Level Lines. The study of a function of two independent variables can be reduced by various means to the study of a function of a single independent variable.

I. We investigate e.g. the function $z = f(x, y)$ "at the point" $P_0(x_0, y_0)$, i.e. in some neighbourhood of this point.

We draw a straight line in any direction through the point $P_0(x_0, y_0)$ in the Oxy plane; its equation will be

$$x - x_0 = \varrho \cos \alpha, \quad y - y_0 = \varrho \sin \alpha, \qquad (*)$$

where ϱ ($\varrho > 0$) is the distance of the point $P(x, y)$ from $P_0(x_0, y_0)$, whilst α is the angle between the straight line and the positive direction of Ox. The value $\varrho = 0$ corresponds to P_0.

We consider the function $f(x, y)$ when its argument — the point $P(x, y)$ — is displaced along the straight line. Since now $x = x_0 + \varrho \cos\alpha$, $y = y_0 + \varrho \sin \alpha$, where x_0, y_0 and α are constants, the function $z = f(x, y)$ becomes a function of the single independent variable ϱ:

$$z = f(x, y) = f(x_0 + \varrho \cos \alpha, \; y_0 + \varrho \sin \alpha) = \varphi_\alpha(\varrho).$$

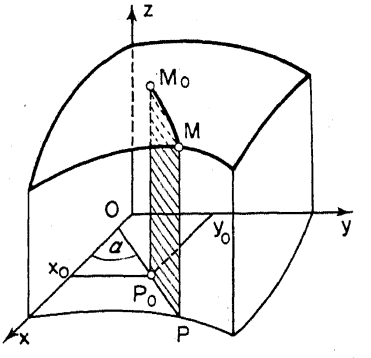

FIG. 11

The graph of this function will be the plane section M_0M of the surface $z = f(x, y)$ by the half-plane passing through the half-line (*) and perpendicular to the Oxy plane (Fig. 11). Knowing the behaviour of $\varphi_\alpha(\varrho)$ at the point $\varrho = 0$ for all possible values of α, we can form an idea of the behaviour of the function $z = f(x, y)$ at the point $P_0(x_0, y_0)$.

II. On assuming that one of the independent variables, say y, is constant: $y = y_0$, we obtain from $z = f(x, y)$ a function of the single independent variable x:

$$z = f(x, y_0),$$

which can be investigated by familiar methods throughout the interval of variation of x of interest to us. Similarly, on setting $x = x_0$, we can consider a function of the single independent variable y:

$$z = f(x_0, y).$$

Knowing the behaviour of these functions of one independent variable for different values of y_0 and x_0, we can describe the vari-

ation of the given function $z = f(x, y)$ throughout the domain of variation of its argument—the point $P(x, y)$.

This method of investigating a function of two variables corresponds geometrically to the investigation of a surface with the aid of sections through it by planes parallel to the co-ordinate planes Oxz and Oyz.

III. A surface can also be investigated with the aid of sections through it by planes parallel to the Oxy plane, i.e. to the plane of the independent variables.

We put $z = z_0$; a relationship is now obtained between the two variables x and y:

$$z_0 = f(x, y).$$

This is the equation of the projection L on to the Oxy plane of the line of intersection of the surface $z = f(x, y)$ with the plane

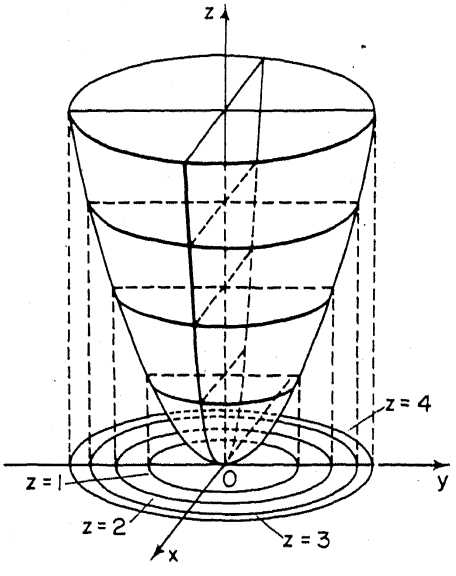

FIG. 12

$z = z_0$. When the point $P(x, y)$ varies along the line L the function maintains a constant value equal to z_0.

Definition. **A *level line* of a function $z = f(x, y)$ is a curve on the Oxy plane at points of which the function preserves a constant value.**

A set of level lines corresponding to the various values of z indicated on the lines is termed a net of curves of the function $z = f(x, y)$. If a net of curves is made for values of z differing only slightly from each other it will show visually the way that $f(x, y)$ behaves for any change of its argument $P(x, y)$.

We give z the values $\ldots -3h, -2h, -h, 0, h, 2h, 3h, \ldots$, where h is any positive number. The net of curves of $f(x, y)$ for these values of z is said to be uniform. (The smaller h, the better the uniform net characterizes the function.) In the parts of the Oxy plane where the level lines of the uniform net are close together the function varies "rapidly", whilst it varies "slowly" in the parts where they are some distance apart. For the first case a "small" and in the second a "large" movement of the point $P(x, y)$ leads to a corresponding variation of the function by the same amount h. Thus the closeness

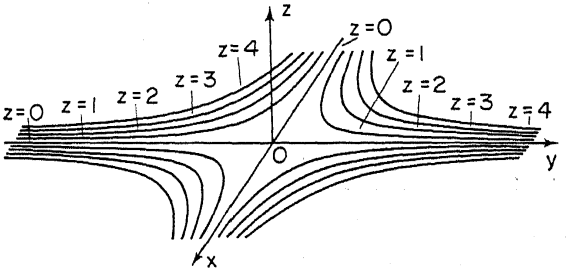

FIG. 13

of the lines of a uniform net can provide an idea of the degree of variation of the function (the amount of rise of the surface $z = f(x, y)$).

Example 1. A uniform net of the function $z = x^2 + y^2$ (paraboloid of revolution) consists of a system of concentric circles with centres at the point $O(0, 0)$ together with point $O(0, 0)$ itself (Fig. 12).

Example 2. The level lines of the function $z = xy$ (hyperbolic paraboloid) consist of equilateral hyperbolas and the Ox and Oy axes (Fig. 13).

We considered in Sec. 72 the Van der Waals equation

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT$$

with different constant temperatures T. Having investigated the relationships between p and v, we drew the relevant graphs and obtained a system of level lines (net of curves) of the function of two variables:

$$T = \frac{1}{R}\left(p + \frac{a}{v^2}\right)(v - b).$$

Example 3. Level lines are often used in the applied sciences for representing a function of two independent variables. For instance, the height above sea level of a point in a locality is regarded as a function of two variables — the co-ordinates of the point — and level lines of the function are drawn on

a map. These are termed "contour lines" in topography. The variation in height in a locality can readily be followed with the aid of a net of "contours". Use is made in meteorology of nets of "isotherms" and "isobars" ("lines of equal temperature" and "lines of equal pressure"), these being the level lines of the temperature and pressure as a function of the point in the atmosphere.

Similarly, "level surfaces" can be defined for functions of three independent variables.

Definition. A surface in space $Oxyz$ at points of which the function $u = f(x, y, z)$ retains a constant value is termed a level surface *of the function.*

The equation of the level surface corresponding to the value $u = u_0$ will be

$$f(x, y, z) = u_0.$$

3. Derivatives and Differentials of Functions of Several Variables

144. Partial Derivatives. Let $z = f(x, y)$ be a function of independent variables x and y which is continuous in some domain. We assign a constant value y_0 to y and consider the function of the single variable x:

$$z = f(x, y_0).$$

Let the function $f(x, y_0)$ be differentiable at the point $x = x_0$, i.e. the limit exists:

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}.$$

We shall write $f'_x(x_0, y_0)$ for this limit, where the subscript x indicates that the derivative is taken with respect to x for fixed y. On assigning different values to x_0 and y_0 (we denote these by x and y), we obtain corresponding values for $f'_x(x, y)$. Consequently $f'_x(x, y)$ is a function of two variables.

The function $f'_x(x, y)$ is termed the partial derivative with respect to x of the function $z = f(x, y)$.

Definition. **The *partial derivative* with respect to x of the function $z = f(x, y)$ is the function of variables x and y obtained by differentiating $f(x, y)$ with respect to x whilst meantime regarding y as a constant.**

The following notations are also used for the partial derivative with respect to x of the function $z = f(x, y)$:

$$\frac{\partial z}{\partial x}, \quad \frac{\partial f(x, y)}{\partial x}, \quad z'_x.$$

The curly ∂ is used here instead of the straight d of the notation for the ordinary derivative, and a subscript is used in addition to the dash to denote the variable x with respect to which the differentiation is carried out. (The dash is sometimes omitted so that we simply write z_x or $f_x(x, y)$.)

The *partial derivative* with respect to y *of the function* $z = f(x, y)$ is defined in exactly the same way:

$$f'_y(x, y) = \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}.$$

It is also written as

$$\frac{\partial z}{\partial y}, \quad \frac{\partial f(x, y)}{\partial y}, \quad z'_y.$$

(We sometimes write z_y, $f_y(x, y)$.)

Since the partial derivative is the ordinary derivative of the function taken on the assumption that only the variable with respect to which the differentiation is carried out varies, the technique of finding partial derivatives (partial differentiation) of elementary functions uses the familiar rules for differentiation of functions of a single variable.

Example 1. We find $\partial z/\partial x$ and $\partial z/\partial y$ for the function $z = 3axy - x^3 - y^3$.

We have, assuming y constant:

$$\frac{\partial z}{\partial x} = 3ay - 3x^2;$$

and assuming x constant:

$$\frac{\partial z}{\partial y} = 3ax - 3y^2.$$

Example 2. We find $\partial z/\partial x$ and $\partial z/\partial y$ of the function $z = x^y$.

When differentiating with respect to x, z is a power function, whilst it is an exponential function when differentiating with respect to y. We obtain:

$$\frac{\partial z}{\partial x} = yx^{y-1}, \quad \frac{\partial z}{\partial y} = x^y \ln x.$$

Example 3. We take the function $r = \sqrt{x^2 + y^2}$. This gives the distance of the point $P(x, y)$ from the origin. We have:

$$r'_x = \frac{\partial r}{\partial x} = \frac{x}{\sqrt{x^2 + y^2}} = \frac{x}{r} = \cos\varphi,$$

$$r'_y = \frac{\partial r}{\partial y} = \frac{y}{\sqrt{x^2 + y^2}} = \frac{y}{r} = \sin\varphi,$$

where φ is the angle between the radius vector of the point $P(x, y)$ and the positive direction of the Ox axis.

The values of the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ at given values $x = x_0$ or $y = y_0$ or $x = x_0$ and $y = y_0$ are written symbolically as

$$\left(\frac{\partial z}{\partial x}\right)_{x=x_0}, \quad \left(\frac{\partial z}{\partial x}\right)_{y=y_0}, \quad \left(\frac{\partial z}{\partial x}\right)_{\substack{x=x_0\\y=y_0}}$$

or

$$f'_x(x_0, y), \quad f'_x(x, y_0), \quad f'_x(x_0, y_0)$$

and so on.

For instance, if $z = 3axy - x^3 - y^3$, we have

$$\left(\frac{\partial z}{\partial x}\right)_{x=a} = 3ay - 3a^2, \quad \left(\frac{\partial z}{\partial x}\right)_{y=a} = 3a^2 - 3x^2, \quad \left(\frac{\partial z}{\partial x}\right)_{\substack{x=a\\y=a}} = 0,$$

$$\left(\frac{\partial z}{\partial y}\right)_{x=a} = 3a^2 - 3y^2, \quad \left(\frac{\partial z}{\partial y}\right)_{y=a} = 3ax - 3a^2, \quad \left(\frac{\partial z}{\partial y}\right)_{\substack{x=a\\y=a}} = 0.$$

The partial derivatives of functions of any number of variables are similarly defined.

Definition. **The *partial derivative of the function* $u = f(x, y, z, \ldots, t)$ *with respect to any one of its arguments* is the function of $x, y, z, \ldots, t$, obtained by differentiating $f(x, y, z, \ldots, t)$ with respect to this argument and meantime assuming that all the remaining arguments are constant.**

For example,

$$\frac{\partial u}{\partial x} = f'_x(x, y, z, \ldots, t)$$

$$= \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y, z, \ldots, t) - f(x, y, z, \ldots, t)}{\Delta x}.$$

Let $r = \sqrt{x^2 + y^2 + z^2}$. This expresses the distance of the point $P(x, y, z)$ from the origin.

We obtain for the partial derivatives:

$$r'_x = \frac{\partial r}{\partial x} = \frac{x}{\sqrt{x^2 + y^2 + z^2}} = \frac{x}{r} = \cos\alpha,$$

$$r'_y = \frac{\partial r}{\partial y} = \frac{y}{r} = \cos\beta;$$

$$r'_z = \frac{\partial r}{\partial z} = \frac{z}{r} = \cos\gamma,$$

where $\cos\alpha, \cos\beta, \cos\gamma$ are the direction cosines of the radius vector of the point $P(x, y, z)$.

The absolute value of the partial derivative $\partial z/\partial x = f'_x(x, y)$ or $\partial z/\partial y = f'_y(x, y)$ gives the rate of change of the function $z = f(x, y)$ when the argument $P(x, y)$ moves along the straight line $y = \text{const}$ or $x = \text{const}$, whilst the sign of the partial derivative f'_x or f'_y indicates the type of change (increase, decrease).

The geometrical interpretation of the partial derivatives is as follows: $f'_x(x_0, y_0)$ is the slope with respect to Ox of the tangent

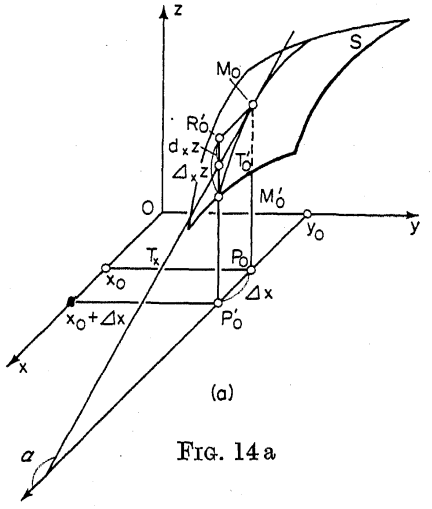

Fig. 14a

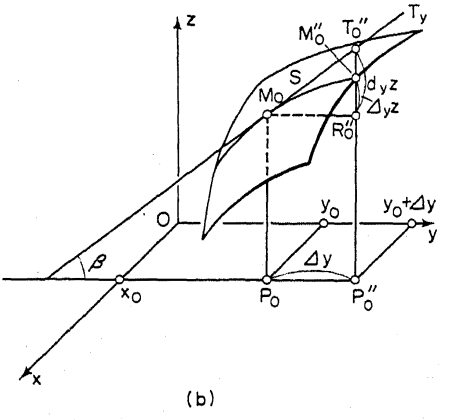

Fig. 14b

at the point $M_0(x_0, y_0, z_0 = f(x_0, y_0))$ to the section of the surface $z = f(x, y)$ by the plane $y = y_0$, i.e. $f'_x(x_0, y_0) = \tan\alpha$ (Fig. 14*a*). It is clear from the figure that in the present case $f'_x(x_0, y_0) < 0$.

The partial derivative $f'_y(x_0, y_0)$ is the slope with respect to Oy of the tangent at the point $M_0(x_0, y_0, z_0)$ to the section of the surface $z = f(x, y)$ by the plane $x = x_0$, i.e. $f'_y(x_0, y_0) = \tan\beta$ (Fig. 14*b*). It is clear from the figure that in the present case $f'_y(x_0, y_0) > 0$.

145. Differentials

I. Partial differentials. The increment that the function $z = f(x, y)$ receives when only one of the variables alters is termed the partial increment of the function with respect to that variable. The following notations are used:

$$\Delta_x z = f(x + \Delta x, y) - f(x, y),$$

$$\Delta_y z = f(x, y + \Delta y) - f(x, y).$$

Definition. ***The partial differential with respect to x of the function $z=f(x, y)$ is the principal part of the increment $f(x+\Delta x, y)-f(x, y)$ proportional to the increment Δx of the independent variable x*** (or, what is just the same, **to the differential dx of this variable**).

The partial differential with respect to y is similarly defined.

The partial differentials are written thus: $d_x z$ is the partial differential with respect to x; $d_y z$ is the partial differential with respect to y.

If the function $z=f(x, y)$ has a partial differential with respect to x at the point $P(x, y)$, it also has at this point a partial derivative $\partial z/\partial x$ and vice versa (see Sec. 51). Now,

$$d_x z = \frac{\partial z}{\partial x}\, dx.$$

Similarly, if the function $z=f(x, y)$ has a partial differential with respect to y at the point $P(x, y)$, it also has a partial derivative $\partial z/\partial y$ at this point and vice versa, where

$$d_y z = \frac{\partial z}{\partial y}\, dy.$$

Thus *the partial differential with respect to either variable of a function of two independent variables is equal to the product of the corresponding partial derivative and the differential of this variable.*

The geometrical meaning of the partial increment $\Delta_x z$ is that it expresses the increment of the z co-ordinate of the surface when the argument of the function—the point P—varies from the position $P_0(x_0, y_0)$ to the position $P_0'(x_0+\Delta x, y_0)$ (Fig. 14a). In the figure, the increment $\Delta_x z<0$; it is represented by the segment $R_0' M_0'$. The partial differential $d_x z$ expresses the increment of the z co-ordinate of the tangent $M_0 T_x$ (Fig. 14a). In our case $d_x z<0$; it is represented by the segment $R_0' T_0'$.

Similarly, the partial increment $\Delta_y z$ expresses the increment of the z co-ordinate of the surface when the argument of a function of the point P moves from the position $P_0(x_0, y_0)$ to the position $P_0''(x_0, y_0+\Delta y)$ (Fig. 14*b*). In the figure, the increment $\Delta_y z>0$; it is represented by the segment $R_0'' M_0''$. The partial differential $d_y z$ expresses the increment of the z co-ordinate of the tangent $M_0 T_y$ (Fig. 14*b*). In our case $d_y z>0$; it is represented by the segment $R_0'' T_0''$.

We find from the formulae for the partial differentials:

$$\frac{\partial z}{\partial x}=\frac{d_x z}{dx}, \quad \frac{\partial z}{\partial y}=\frac{d_y z}{dy}.$$

It is clear from this that the partial derivatives can be regarded, as in the case of the ordinary derivative, as fractions provided the *corresponding* partial differential is written in the numerator of each fraction and the differential of the independent variable in the denominator. On the other hand the symbols $\partial z/\partial x$ and $\partial z/\partial y$ are to be regarded as unique single entities and not as fractions, since even if we agree to let ∂x and ∂y denote dx and dy, ∂z will denote different quantities in the first and second cases ($d_x z$ and $d_y z$).

We take as an example the Mendeleev–Clapeyron equation

$$pv = RT$$

and find from this $\partial p/\partial v$, $\partial v/\partial T$, $\partial T/\partial p$. We have

$$\frac{\partial p}{\partial v} = \frac{\partial}{\partial v}\left(\frac{RT}{v}\right) = -\frac{RT}{v^2},$$

$$\frac{\partial v}{\partial T} = \frac{\partial}{\partial T}\left(\frac{RT}{p}\right) = \frac{R}{p},$$

$$\frac{\partial T}{\partial p} = \frac{\partial}{\partial p}\left(\frac{pv}{R}\right) = \frac{v}{R}.$$

The product of these three partial derivatives yields a relationship of importance in thermodynamics:

$$\frac{\partial p}{\partial v}\,\frac{\partial v}{\partial T}\,\frac{\partial T}{\partial p} = -\frac{RT}{v^2}\,\frac{R}{p}\,\frac{v}{R} = -\frac{RT}{pv} = -1.$$

If the symbols of the partial derivatives in terms of curly ∂ were in fact fractions, we should obtain 1 instead of -1 for the product on the left-hand side.

Partial increments and partial differentials are defined for functions of any number of independent variables in the same way as for functions of two variables.

Definition. **The *partial differential of the function* $u = f(x, y, z, \ldots, t)$ *with respect to any one of its arguments* is the principal part of the corresponding partial increment, proportional to the increment (differential) of the independent variable.**

It follows readily from the definition of partial derivative, as above, that

$$d_x u = \frac{\partial u}{\partial x}\,dx, \quad d_y u = \frac{\partial u}{\partial y}\,dy, \ldots$$

Consequently, the partial differential of a function of several independent variables with respect to one of them is equal to the corresponding partial derivative multiplied by the differential of the variable concerned.

II. TOTAL DIFFERENTIAL. Let the function $z = f(x, y)$ be continuous and differentiable with respect to x and y; we can now find, with the aid of the partial differentials, expressions as accurate as may be desired for the increments of the function for sufficiently small displacements of the point $P(x, y)$ in directions parallel to Ox and Oy. It is natural to look for an expression for the increment of the function $z = f(P) = f(x, y)$ for an *arbitrary* displacement of its argument $P(x, y)$ (not only in directions parallel to Ox and Oy). The increment

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y)$$

for arbitrary Δx and Δy is termed the total increment of the function $z = f(x, y)$ at the point $P(x, y)$.

The expression for the total increment of a function in terms of arbitrary increments of the independent variables is extremely complicated; there is only one case in which the expression is simple, namely when the function $f(x, y)$ is linear: $f(x, y) = ax + by + c$; here, as may easily be seen,

$$\Delta z = a\,\Delta x + b\,\Delta y.$$

It happens, however, (see Sec. 51) that constant coefficients a and b can usually be chosen for a given point $P(x, y)$ such that the expression $a\,\Delta x + b\,\Delta y$, whilst not strictly equal to Δz, only differs from Δz by a higher order infinitesimal than Δx and Δy (assuming*

* We assume in addition that Δx and Δy are infinitesimals of the same order. It is now easily seen that $\varrho = \sqrt{\Delta x^2 + \Delta y^2}$ is also of the same order, this being the infinitesimal displacement of the argument of the function, i.e. of the point $P(x, y)$. We have, in fact:

$$\frac{\Delta x}{\sqrt{\Delta x^2 + \Delta y^2}} = \frac{1}{\sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}} \to \frac{1}{\sqrt{1 + k^2}} \neq 0,$$

$$\frac{\Delta y}{\sqrt{\Delta x^2 + \Delta y^2}} = \frac{k}{\sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}} \to \frac{k}{\sqrt{1 + k^2}} \neq 0,$$

where

$$\lim \frac{\Delta y}{\Delta x} = k \neq 0.$$

that Δx, Δy, and therefore Δz also are infinitesimals):

$$\Delta z = a\Delta x + b\Delta y + \alpha, \qquad (*)$$

where

$$\lim_{\Delta x \to 0} \frac{\alpha}{\Delta x} = 0 \quad \text{and} \quad \lim_{\Delta y \to 0} \frac{\alpha}{\Delta y} = 0,$$

or, what amounts to the same thing,

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\alpha}{\sqrt{\Delta x^2 + \Delta y^2}} = 0, \quad \text{i.e.} \quad \lim_{\varrho \to 0} \frac{\alpha}{\varrho} = 0.$$

The sum $a\Delta x + b\Delta y$ is termed the differential, or sometimes the total differential, to distinguish it from the partial differentials, of the function $z = f(x, y)$ at the point $P(x, y)$; it is written as dz or $df(x, y)$:

$$dz = adx + bdy \qquad (**)$$

(as previously, $\Delta x = dx$, $\Delta y = dy$).

We compare Δz and dz. If $a = b = 0$, the differential dz is equal to zero and cannot be equated to any other infinitesimal, including Δz. With $a \neq 0$ or $b \neq 0$, Δz and dz are equivalent infinitesimals, i.e. in other words, dz is the principal part of Δz (see Sec. 39). We have, in fact:

$$\frac{\Delta z}{dz} = \frac{dz + \alpha}{dz} = 1 + \frac{\alpha}{dz},$$

and since $\alpha/dz = \alpha/(a\Delta x + b\Delta y) \to 0$, $\Delta z/dz \to 1$. We can therefore say that dz is the principal part of Δz (assuming that $\Delta x \to 0$, $\Delta y \to 0$), which is either linear with respect to Δx and Δy, or zero.

In the definition of the differential we cannot provide for what is in fact the very exceptional case when $a = b = 0$. (In this case $dz = 0$ and Δz is itself an infinitesimal of higher order than Δx and Δy.)

Definition. **The *(total) differential* of a function of two independent variables is the principal part of the (total) increment of the function, linear in the increments of the independent variables.**

Let the function $z = f(x, y)$ have a differential at the point $P(x, y)$, i.e. we can extract from the increment of the function $\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y)$ a "principal part" which is linear in Δx and Δy, i.e. we can write equation (*).

THEOREM. **The (total) differential of a function of two independent variables is equal to the sum of the products of the partial derivatives of the function and the differentials of the corresponding independent variables.**

Proof. Equation (**) for the differential holds for any dx and dy, i.e. in particular for $dy = 0$. In this case $\Delta z = \Delta_x z$, and we get

$$d_x z = a\,dx,$$

whence

$$a = \frac{d_x z}{dx} = f'_x(x, y).$$

It may be shown in the same way that $b = f'_y(x, y)$.

The expression for the differential at an arbitrary point $P(x, y)$ reads:

$$dz = f'_x(x, y)\,dx + f'_y(x, y)\,dy$$

or

$$dz = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy.$$

This is what we wanted to show.

Example 1. Let $z = 3axy - x^3 - y^3$. Then

$$dz = (3ay - 3x^2)\,dx + (3ax - 3y^2)\,dy.$$

Example 2. We have for the function $z = x^y$:

$$dz = y\,x^{y-1}\,dx + x^y \ln x\,dy.$$

Example 3. If $r = \sqrt{x^2 + y^2}$, we have

$$dr = \cos\varphi\,dx + \sin\varphi\,dy.$$

Since

$$\frac{\partial z}{\partial x}\,dx = d_x z \quad \text{and} \quad \frac{\partial z}{\partial y}\,dy = d_y z,$$

we have

$$dz = d_x z + d_y z,$$

i.e. *the differential of a function of two independent variables is equal to the sum of its partial differentials.*

Thus the principal part of the increment of $z = f(P)$ when the point P is displaced in an *arbitrary* direction is equal to the sum of the principal parts* of the increments obtained when point P is displaced *along the co-ordinate axes.*

* The principal parts are linear in the increments of the independent variables.

This is the exact mathematical expression of the so-called principle of superposition of small operations, which is often used in the natural sciences. It can be stated briefly as:

The simultaneous result of two changes (which are sufficiently small) *is given* to any required accuracy *by the sum of the results of each change separately.*

If the total differential dz of the function exists at a point $P(x, y)$, we obtain, on taking this instead of the true increment Δz, an approximate expression with "unlimited accuracy". This means that $dz = d_x z + d_y z$ is approximately equal to Δz in a sufficiently small neighbourhood of the point $P(x, y)$ (i.e. with sufficiently small Δx and Δy), with a relative error of any required smallness. At the same time we preserve the simplicity of the expression for the increment of the function (viz. linearity in Δx and Δy), which holds strictly only for a linear function.

The differential of a function is thus easily found from the values of the partial derivatives at the initial point, i.e. from $\partial z/\partial x$ and $\partial z/\partial y$, and from the displacements of the arguments of the function in the directions of the axes (i.e. from Δx and Δy).

Definition. A function of two independent variables which has a differential at a given point is said to be differentiable at this point.

The definition of differential may be carried over to functions of any number of independent variables.

Definition. **The *(total) differential* du of a function of several independent variables $u = f(x, y, z, \ldots, t)$ is the principal part of the (total) increment**

$$\Delta z = f(x + \Delta x, y + \Delta y, z + \Delta z, \ldots, t + \Delta t) - f(x, y, z, \ldots, t),$$

which is linear in the increments of the independent variables Δx, Δy, Δz, ..., Δt.

We can prove a theorem similar to the above.

THEOREM. **If a function u has a total differential du, then**

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy + \frac{\partial u}{\partial z}\,dz + \cdots + \frac{\partial u}{\partial t}\,dt,$$

or

$$du = d_x u + d_y u + d_z u + \cdots + d_t u,$$

i.e. the differential of a function of several variables is equal to the sum of its partial differentials.

The connection between the increment and differential is given by

$$\Delta u = du + \alpha,$$

where α is an infinitesimal of higher order than the distance

$$\varrho = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2 + \cdots + \Delta t^2},$$

by which the point $P(x, y, z, \ldots, t)$—the argument of the function—is displaced.

A function $f(x, y, z, \ldots, t)$ is said to be differentiable at a point $P(x, y, z, \ldots, t)$ if it has a differential at this point.

REMARK. If $du = 0$, u is a constant. For, it follows from the identity $\frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy + \cdots = 0$ that $\frac{\partial u}{\partial x} = 0$, $\frac{\partial u}{\partial y} = 0$, ... identically, i.e. that u is independent of $x, y, z, \ldots$ i.e. is constant.

146. Geometrical Interpretation of the Differential. Just as the derivative and differential of a function of one variable are connected with the tangent to a curve—the graph of the function—, the derivatives and differential of a function of two variables are connected with the tangent plane to a surface—the graph in this case.

Let the function $z = f(x, y)$ be differentiable at the point $P_0(x_0, y_0)$. We consider the sections by the planes $y = y_0$ and $x = x_0$ of the surface S representing this function. We draw tangents M_0T_x and M_0T_y at the point $M_0(x_0, y_0, z_0)$ to the plane curves thus obtained on the surface (Fig. 15). These two straight lines intersecting at the point M_0 define a plane T which is called the tangent plane to the surface S at the point M_0. The point M_0 is called the point of contact of the tangent plane T with the surface S. Let us find the equation of the tangent plane. The straight line M_0T_x lies in the plane $y = y_0$, parallel to the Oxy plane, its slope with respect to Ox being $f'_x(x_0, y_0)$. The equations of the straight line M_0T_x are therefore:

$$z - z_0 = f'_x(x_0, y_0)\,(x - x_0), \quad y = y_0.$$

The equations of the straight line M_0T_x are similarly found as:

$$z - z_0 = f'_y(x_0, y_0)\,(y - y_0), \quad x = x_0.$$

Since the plane T passes through the point $M_0(x_0, y_0, z_0)$, its equation can be written as

$$z - z_0 = A(x - x_0) + B(y - y_0).$$

The straight lines M_0T_x and M_0T_y lie in plane T; their equations must thus coincide with the equation of the plane. On substituting

in this latter expression for $z - z_0$ and $y - y_0$ from the equations of M_0T_x, we get

$$f'_x(x_0, y_0)\,(x - x_0) = A(x - x_0),$$

whence

$$A = f'_x(x_0, y_0).$$

Similarly we find that

$$B = f'_y(x_0, y_0).$$

The required equation of the tangent plane is thus

$$z - z_0 = f'_x(x_0, y_0)\,(x - x_0) + f'_y(x_0, y_0)\,(y - y_0).$$

We shall show in Sec. 166 that this plane contains the tangent at the point $M_0(x_0, y_0, z_0)$ to any curve on the surface S passing through the point $M_0(x_0, y_0, z_0)$.

The equation of the tangent plane may be written more briefly as

$$z - z_0 = \frac{\partial z}{\partial x}(x - x_0) + \frac{\partial z}{\partial y}(y - y_0), \qquad (*)$$

though it must be borne in mind here that the coefficients of $x - x_0$ and $y - y_0$ are the *values* of the partial derivatives in question at the point $P_0(x_0, y_0)$.

The geometrical meaning of the differential of a function of two independent variables follows from the following proposition.

THEOREM. **The differential of the function $z = f(x, y)$ at the point $P_0(x_0, y_0)$ is represented by the increment of the z co-ordinate of the tangent plane to the surface $z = f(x, y)$ at the corresponding point $M_0(x_0, y_0, z_0)$ of the surface.**

Proof. The right-hand side of the equation of the tangent plane (*) is in fact the expression for the differential of the function $z = f(x, y)$. In view of this we can write the equation of the tangent plane in the form

$$z - z_0 = (dz)_{P_0};$$

here z_0 is the z co-ordinate of the point of contact, z is the current z co-ordinate of the plane, and $(dz)_{P_0}$ is the differential of the function $z = f(x, y)$ evaluated at the point $P_0(x_0, y_0)$ corresponding to the point of contact $M_0(x_0, y_0, z_0)$. This is what we wished to prove.

Let the point $P(x, y)$—the argument of the function $z = f(x, y)$—be displaced from the position $P_0(x_0, y_0)$ to the position $P_1(x_0 + \Delta x, y_0 + \Delta y)$ (Fig. 15). The increment Δz is now represented by the segment R_1M_1—the increment of the z co-ordinate of the

surface S, whilst the differential dz is given by R_1T_1—the increment of the z co-ordinate of the tangent plane T.

In the particular case when the point P moves from $P_0(x_0, y_0)$ to $P_0'(x_0 + \Delta x, y_0)$, the differential dz reduces to the partial differential $d_x z$ and is given by $R_0'T_0'$ (the point T_0' lies on the straight line M_0T_x). In the other particular case, when P moves from

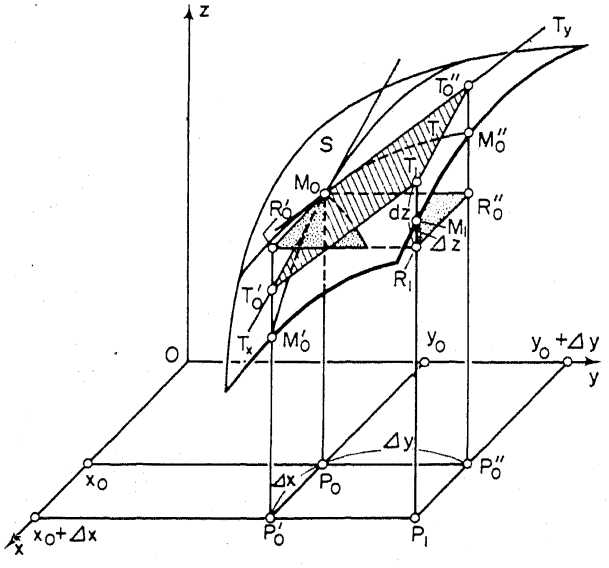

FIG. 15

$P_0(x_0, y_0)$ to $P_0''(x_0, y_0 + \Delta y)$, the differential dz becomes the partial differential $d_y z$ and is given by $R_0''T_0''$ (the point T_0'' lies on the straight line M_0T_y).

The deviation of the differential from the increment of the function, i.e. the difference $dz - \Delta z$, is represented by the segment M_1T_1 lying between the surface S and the tangent plane T. We can say that $dz - \Delta z$ measures the distance from the surface to the tangent plane with respect to the z axis. It will be seen that this distance is an infinitesimal of higher order than the distance $\varrho = P_0P_1$.

147. Application of the Differential to Approximations. If we put $\Delta z \approx dz$ for points $P(x, y)$ of some neighbourhood of the point $P_0(x_0, y_0)$, i.e. we neglect the term α in the right-hand side of the strict equation

$$f(x, y) - f(x_0, y_0) = f_x'(x_0, y_0)\,(x - x_0) + f_y'(x_0, y_0)\,(y - y_0) + \alpha,$$

we obtain the approximate equation

$$f(x, y) - f(x_0, y_0) \approx f'_x(x_0, y_0)\,(x - x_0) + f'_y(x_0, y_0)\,(y - y_0),$$

or

$$f(x, y) \approx f(x_0, y_0) + f'_x(x_0, y_0)\,(x - x_0) + f'_y(x_0, y_0)\,(y - y_0), \quad (*)$$

expressing the given function as a *linear* function of the independent variables. (The error of approximate equation (*) will be found with the aid of Taylor's formula for a function of two independent variables (see Sec. 155).)

Geometrically, the substitution of formula (*) for the given function $f(x, y)$ in the neighbourhood of the point $P_0(x_0, y_0)$ implies replacing a piece of the surface $z = f(x, y)$ by the corresponding piece of the tangent plane to the surface at the point $M_0(x_0, y_0, z_0) = f(x_0, y_0)$. Over small areas such a substitution leads, as may be seen, to a small *relative* error in finding the values of the function (i.e. the z co-ordinate of the surface).

The approximate equation (*) is used in practice primarily for solving problems of the two types below.

I. *Given the values of* $f(x_0, y_0)$, $f'_x(x_0, y_0)$, $f'_y(x_0, y_0)$, Δx, Δy, *to find the approximate value of* $f(x_0 + \Delta x, y_0 + \Delta y)$.

We have from expression (*):

$$f(x_0 + \Delta x, y_0 + \Delta y) \approx f(x_0, y_0) + f'_x(x_0, y_0)\,\Delta x + f'_y(x_0, y_0)\,\Delta y.$$

The following examples are for illustration:

Example 1. The hypotenuse c and acute angle α are varied simultaneously in a right-angled triangle; knowing the adjacent sides

$$a = c \sin \alpha, \quad b = c \cos \alpha$$

for certain values of c and α, we can find the adjacent sides a_1 and b_1 for neighbouring values $c + \Delta c, \alpha + \Delta\alpha$. Assuming Δc and $\Delta\alpha$ to be small, we replace increments Δa and Δb by the differentials da and db; now,

$$a_1 \approx a + da, \quad b_1 \approx b + db.$$

Since

$$da = \sin\alpha\,\Delta c + c \cos\alpha\,\Delta\alpha, \quad db = \cos\alpha\,\Delta c - c \sin\alpha\,\Delta\alpha,$$

we have

$$a_1 \approx a + \sin\alpha\,\Delta c + c \cos\alpha\,\Delta\alpha, \quad b_1 \approx b + \cos\alpha\,\Delta c - c \sin\alpha\,\Delta\alpha.$$

For example, let $c = 2$, $\alpha = 30°$; we find sides a_1 and b_1 for $c_1 = 2{\cdot}1$, $\alpha_1 = 31°$.

We have:

$$a_1 \approx 2 \cdot \frac{1}{2} + \frac{1}{2} \cdot 0{\cdot}1 + 2 \cdot \frac{\sqrt{3}}{2} \cdot \frac{\pi}{180},$$

$$b_1 \approx 2 \cdot \frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2} \cdot 0{\cdot}1 - 2 \cdot \frac{1}{2} \cdot \frac{\pi}{180},$$

i.e. $$a_1 \approx 1{\cdot}080, \quad b_1 \approx 1{\cdot}801.$$

This result may be verified by direct working (from the formulae $a_1 = c_1 \sin \alpha_1$, $b_1 = c_1 \cos \alpha_1$, see Sec. 53).

Example 2. The side a in a triangle with angles α, β, γ and opposite sides a, b, c can be found with the aid of the formula

$$a = \sqrt{b^2 + c^2 - 2bc \cos \alpha}.$$

Let sides b and c and angle α be given small increments Δb, Δc and $\Delta \alpha$ respectively. Putting $\Delta a \approx da$, we have from the formula for the differential of a function of three variables:

$$\Delta a \approx \frac{b - c \cos \alpha}{a} \Delta b + \frac{c - b \cos \alpha}{a} \Delta c + \frac{bc \sin \alpha}{a} \Delta \alpha;$$

but it is easily seen that $b - c \cos \alpha = a \cos \gamma$, $c - b \cos \alpha = a \cos \beta$, so that

$$\Delta a \approx \cos \gamma \, \Delta b + \cos \beta \, \Delta c + \frac{bc}{a} \sin \alpha \, \Delta \alpha.$$

This formula enables us to find the increment received by side a, given the values of the other sides b, c, the angle α, and the increments Δb, Δc and $\Delta \alpha$ which show how the latter vary.

II. *The values of* $f(x_0, y_0)$, $f'_x(x_0, y_0)$, $f'_y(x_0, y_0)$ *are known; given errors* δ' *and* δ'' *in the values of* x_0 *and* y_0 ($|\Delta x| < \delta'$, $|\Delta y| < \delta''$), *to find the error* ε *when the value* $f(x_0, y_0)$ *is taken as an approximation to* $f(x_0 + \Delta x, y_0 + \Delta y)$.

We have here:

$$|\Delta z| \leqslant |f'_x(x_0, y_0)| \cdot |\Delta x| + |f'_y(x_0, y_0)| \cdot |\Delta y| \leqslant |f'_x(x_0, y_0)| \, \delta' +$$
$$+ |f'_y(x_0, y_0)| \, \delta'' \leqslant (|f'_x(x_0, y_0)| + |f'_y(x_0, y_0)|) \, \bar{\delta} = \varepsilon,$$

where $\bar{\delta}$ is the greater of the numbers δ' and δ''.

If we take $f(x_0, y_0)$ instead of the accurate value $f(x_0 + \Delta x, y_0 + \Delta y)$, the error involved is ε as just mentioned. We can work out from this what the error $\bar{\delta}$ must be in order for a previously assigned value of the error ε in $f(x_0, y_0)$ not to be exceeded:

$$\bar{\delta} = \frac{\varepsilon}{|f'_x(x_0, y_0)| + |f'_y(x_0, y_0)|}.$$

A formula for the relative error is easily obtained from the above.

We shall take as examples approximate evaluations of products and quotients.

Example 1. Let $z = xy$, $z_0 = x_0 y_0$. Now with small Δx and Δy:

$$|\Delta z| \leqslant |y_0||\Delta x| + |x_0||\Delta y|,$$

whence

$$\left|\frac{\Delta z}{z_0}\right| \leqslant \left|\frac{\Delta x}{x_0}\right| + \left|\frac{\Delta y}{y_0}\right|,$$

i.e. *the maximum relative error in the product is equal to the sum of the relative errors of the factors.*

Example 2. If $z = x/y$, $z_0 = x_0/y_0$, we find similarly that

$$|\Delta z| \leqslant \left|\frac{\Delta x}{y_0}\right| + \left|\frac{x_0}{y_0^2}\right||\Delta y|,$$

whence

$$\left|\frac{\Delta z}{z_0}\right| \leqslant \left|\frac{\Delta x}{x_0}\right| + \left|\frac{\Delta y}{y_0}\right|.$$

i.e. *the maximum relative error of the quotient is equal to the sum of the relative errors of top and bottom.*

148. Directional Derivatives. I. Let the argument of the function $f(x, y)$—the point $P(x, y)$—vary along a given radius vector, drawn from the point P_0 and forming an angle α with the positive direction of Ox. We have (Sec. 143):

$$\left.\begin{aligned} x - x_0 &= \varrho \cos\alpha, \\ y - y_0 &= \varrho \sin\alpha = \varrho \cos\beta \left(\beta = \frac{\pi}{2} - \alpha\right) \end{aligned}\right\} \qquad (*)$$

and

$$f(P) = f(x_0 + \varrho\cos\alpha,\ y_0 + \varrho\sin\alpha).$$

We form the ratio

$$\frac{f(P) - f(P_0)}{PP_0} = \frac{f(x_0 + \varrho\cos\alpha,\ y_0 + \varrho\sin\alpha) - f(x_0, y_0)}{\varrho}. \qquad (**)$$

This is the ratio of the increment $\Delta z = f(P) - f(P_0)$ of the function $z = f(P)$ as a function of the single variable ϱ to the increment of this variable (since the value $\varrho = 0$ corresponds to P_0).

Now let point P tend to point P_0 along the radius vector. Angle α will now remain constant, whilst $\varrho \to 0$.

Definition. **If the ratio (**) has a limit as $\varrho \to 0$, it is termed the *derivative* of the function $z = f(x, y)$ with respect to the *direction* α at the point $P_0(x_0, y_0)$.**

This limit is denoted by $f'_\alpha(x_0, y_0)$ or $(\partial z/\partial \alpha)_{P_0}$; any other indication of the required direction can be written instead of α.

THEOREM. **If the function $z = f(x, y)$ is differentiable at the point $P_0(x_0, y_0)$, it has a derivative $f'_\alpha(x_0, y_0)$ at this point with respect to any direction α, where**

$$f'_\alpha(x_0, y_0) = f'_x(x_0, y_0) \cos \alpha + f'_y(x_0, y_0) \sin \alpha.$$

Proof. We write down the formula connecting the increment Δz and the differential dz corresponding to a shift of the argument from the point $P_0(x_0, y_0)$ to $P(x, y)$:

$$f(P) - f(P_0) = f(x, y) - f(x_0, y_0)$$
$$= f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) + \varepsilon,$$

or

$$f(x_0 + \varrho \cos \alpha, y_0 + \varrho \sin \alpha) - f(x_0, y_0)$$
$$= f'_x(x_0, y_0)\, \varrho \cos\alpha + f'_y(x_0, y_0)\, \varrho \sin\alpha + \varepsilon,$$

where ε is an infinitesimal of higher order than ϱ. Hence

$$\frac{f(x_0 + \varrho \cos \alpha, y_0 + \varrho \sin \alpha) - f(x_0, y_0)}{\varrho}$$
$$= f'_x(x_0, y_0) \cos \alpha + f'_y(x_0, y_0) \sin \alpha + \frac{\varepsilon}{\varrho}.$$

Since $\varepsilon/\varrho \to 0$ as $\varrho \to 0$, the limit of the ratio on the left-hand side exists and is equal to

$$f'_x(x_0, y_0) \cos \alpha + f'_y(x_0, y_0) \sin \alpha.$$

Consequently, given the differentiability of $z = f(x, y)$ at $P_0(x_0, y_0)$, we have

$$f'_\alpha(x_0, y_0) = f'_x(x_0, y_0) \cos \alpha + f'_y(x_0, y_0) \sin \alpha.$$

This is what we had to prove.

If the point $P_0(x_0, y_0)$ is fixed, the derivative with respect to direction α is a function of α only $(0 \leqslant \alpha < 2\pi)$.

We have in the particular cases when $\alpha = 0$ and $\alpha = \frac{1}{2}\pi$:

$$f'_0(x, y) = \frac{\partial z}{\partial x} \quad \text{and} \quad f'_{\frac{\pi}{2}}(x, y) = \frac{\partial z}{\partial y}.$$

Therefore the *partial derivatives* $\partial z/\partial x$ *and* $\partial z/\partial y$ *are the derivatives with respect to the positive directions of the* Ox *and* Oy *axes.*

The absolute value of the directional derivative $\partial z/\partial \alpha = f'_\alpha(x, y)$ indicates the rate of change of function $z = f(x, y)$ when the point $P(x, y)$ moves along a given direction $\alpha = \text{const}$, whilst its sign shows the nature of the variation (increase, decrease).

It may be observed that the derivative with respect to a given direction α is equal to the derivative in the opposite direction α' taken with the opposite sign. For $\alpha' = \pi + \alpha$, i.e.

$$\frac{\partial z}{\partial \alpha'} = \frac{\partial z}{\partial x}\cos(\pi + \alpha) + \frac{\partial z}{\partial y}\sin(\pi + \alpha)$$

$$= -\frac{\partial z}{\partial x}\cos\alpha - \frac{\partial z}{\partial y}\sin\alpha = -\frac{\partial z}{\partial \alpha}.$$

II. *Definition. The* directional derivative *of a given "smooth" (i.e. having a continuously varying tangent) curve* L *at a point* P_0 *is defined as the derivative with respect to the direction of the tangent to* L *at* P_0. This derivative is often written as $\partial z/\partial s$.

Let the function $z = f(P)$ be differentiable in a neighbourhood of the point P_0. We shall show that a point on the curve can be taken instead of the point P' on the tangent when finding $\partial z/\partial s$ as the limit of the ratio $\dfrac{f(P') - f(P_0)}{P'P_0}$.

We have, in fact:

$$\frac{f(P') - f(P_0)}{P'P_0} = \left[\frac{f(P) - f(P_0)}{PP_0} + \frac{f(P') - f(P)}{P'P}\,\frac{P'P}{PP_0}\right]\frac{PP_0}{P'P_0},$$

where P is a point on L having the same abscissa as P'. When $P' \to P_0$, the point P also tends to P_0 and $PP_0/P'P_0 \to 1$, $P'P/PP_0 \to 0$, whilst the ratio $[f(P') - f(P)]/P'P$ remains bounded, inasmuch as the function $f(P)$ is differentiable.

Therefore

$$\lim_{P' \to P_0} \frac{f(P') - f(P_0)}{P'P_0} = \lim_{P \to P_0} \frac{f(P) - f(P_0)}{PP_0}.$$

In particular, if the curve L is a level line of the function $z = f(P)$, we have $f(P) = f(P_0)$ when P belongs to L, so that $\partial z/\partial s = 0$.

Thus *the derivative of a function* $z = f(P)$ *with respect to one of its level lines is equal to zero.*

This can be taken as characterizing a level line as a line of constant values of the function.

III. We shall take some examples of directional differentiation.

Example 1. We take the function $r = \sqrt{x^2 + y^2}$ (the radius vector of the point $P(x, y)$). Since $r'_x = \cos\varphi$, $r'_y = \sin\varphi$, where φ is the angle between radius vector r and the positive direction of Ox (see Sec. 14), we have

$$\frac{\partial r}{\partial \alpha} = r'_x \cos\alpha + r'_y \sin\alpha = \cos\varphi \cos\alpha + \sin\varphi \sin\alpha = \cos(\varphi - \alpha).$$

In particular, *the derivative of the radius vector with respect to its own direction ($\alpha = \varphi$) is always equal to unity, whilst it is always zero with respect to the perpendicular direction.* This has a simple meaning: the radius vector changes uniformly with respect to its own direction, with a rate equal to unity, whilst it does not change at all in a direction perpendicular to it.

Example 2. The derivative is sometimes taken of the function $z = f(x, y)$ with respect to the direction of the radius vector r of the point $P(x, y)$ (i.e. with $\alpha = \varphi$). It is usually written as $\partial z/\partial r$:

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\cos\varphi + \frac{\partial z}{\partial y}\sin\varphi.$$

In particular, for $z = x$ and $z = y$ we have

$$\frac{\partial x}{\partial r} = 1 \cdot \cos\varphi + 0 \cdot \sin\varphi = \cos\varphi$$

and

$$\frac{\partial y}{\partial r} = 0 \cdot \cos\varphi + 1 \cdot \sin\varphi = \sin\varphi.$$

We can therefore write

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r}.$$

It may be shown by the reader that, for any direction α,

$$\frac{\partial z}{\partial \alpha} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial \alpha} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial \alpha}.$$

IV. The definition of directional derivative for a function of three independent variables is similar to the above:

$$u = f(P) = f(x, y, z).$$

We take a point $P_0(x_0, y_0, z_0)$ and any direction P_0N originating from P_0. Let the direction cosines of P_0N be $\cos\alpha$, $\cos\beta$, $\cos\gamma$, and

let $P(x, y, z)$ be a point on the half line P_0N distinct from P_0, ϱ being the distance of P from P_0. Then

$$x = x_0 + \varrho \cos\alpha, \quad y = y_0 + \varrho \cos\beta, \quad z = z_0 + \varrho \cos\gamma.$$

Definition. **The limit**

$$\lim_{\varrho \to 0} \frac{f(P) - f(P_0)}{\varrho}$$

$$= \lim_{\varrho \to 0} \frac{f(x_0 + \varrho\cos\alpha,\ y_0 + \varrho\cos\beta,\ z_0 + \varrho\cos\gamma) - f(x_0, y_0, z_0)}{\varrho},$$

if it exists, is called the *derivative of the function* $u = f(P)$ *with respect to the direction* P_0N *at the point* P_0.

We shall write $f'_{P_0N}(P_0)$ or $f'_{P_0N}(x_0, y_0, z_0)$ for the derivative with respect to the direction P_0N.

THEOREM. **If a function $u = f(x, y, z)$ is differentiable at the point $P_0(x_0, y_0, z_0)$, it must have a derivative $f'_{P_0N}(x_0, y_0, z_0)$ at this point with respect to any direction P_0N, whilst**

$$f'_{P_0N}(x_0, y_0, z_0) = f'_x(x_0, y_0, z_0)\cos\alpha + f'_y(x_0, y_0, z_0)\cos\beta + \\ + f'_z(x_0, y_0, z_0)\cos\gamma.$$

The proof is just the same as in the case of two independent variables.

In particular, we find when $\alpha = 0$ ($\beta = 0, \gamma = 0$) that the derivative with respect to the positive direction of Ox (Oy, Oz) is the partial derivative $\partial u/\partial x$ ($\partial u/\partial y$, $\partial u/\partial z$).

It can be shown as above that the *derivative of a function $u = f(x, y, z)$ with respect to any direction tangential to a level surface of the function is equal to zero.*

149. Differentiability of Functions of Two Independent Variables. We described a continuous function of one independent variable $y = f(x)$ as differentiable at a point $P_0(x_0)$ if the function has a differential at this point. This proved to be equivalent (see Sec. 51) to the condition that $y = f(x)$ has a derivative at the point $P_0(x_0)$. The matter is more complicated for functions of two independent variables.

We have already described a function of two independent variables $z = f(x, y)$ as differentiable at a point $P_0(x_0, y_0)$ (Sec. 145) if it has a *differential* at this point. This is no longer equivalent, however, to the existence of derivatives of $f(x, y)$ at the point $P_0(x_0, y_0)$.

It may be seen from examples (see below) that the existence of the partial derivatives is in fact insufficient by itself to ensure that a principal part linear in Δx and Δy can be extracted from Δz. Hence the differentiability of $z = f(x, y)$ with respect to each of its arguments (i.e. the existence at a given point of the partial differentials $d_x z = f'_x dx$ and $d_y z = f'_y dy$) does not imply the differentiability of $f(x, y)$ as a function of an arbitrarily varying point $P(x, y)$ (i.e. the existence of a total differential dz). On the other hand, if dz exists, $d_x z$ and $d_y z$ must also exist and $dz = d_x z + d_y z$.

Furthermore, even the existence at a point $P_0(x_0, y_0)$ of derivatives of $z = f(x, y)$ with respect to any direction does not imply the existence of a differential of the function. We may take as an example $z = f(x, y) = \sqrt[3]{x^3 + y^3}$ and consider it at the point $P_0(0, 0)$.

We find the derivative with respect to a direction α. We have:

$$\frac{\Delta z}{\varrho} = \frac{f(0 + \Delta x, 0 + \Delta y) - f(0, 0)}{\varrho} = \frac{\sqrt[3]{\Delta x^3 + \Delta y^3}}{\varrho},$$

or

$$\varrho = \sqrt{\Delta x^2 + \Delta y^2},$$

$$\frac{\Delta z}{\varrho} = \frac{\sqrt[3]{(\varrho \cos \alpha)^3 + (\varrho \sin \alpha)^3}}{\varrho} = \sqrt[3]{\cos^3 \alpha + \sin^3 \alpha},$$

whence we find, as $\varrho \to 0$:

$$f'_\alpha(0, 0) = \sqrt[3]{\cos^3 \alpha + \sin^3 \alpha};$$

in particular, $f'_x(0, 0) = 1$ and $f'_y(0, 0) = 1$.

At the same time $f(x, y)$ does not have a differential at the origin. In fact, if a differential dz were to exist, it would be equal to $f'_x(0, 0)\,dx + f'_y(0, 0)\,dy = dx + dy$, and the difference $\Delta z - (dx + dy)$ must be an infinitesimal of higher order than $\varrho = \sqrt{dx^2 + dy^2}$. But

$$\Delta z - (dx + dy) = \varrho \sqrt[3]{\cos^3 \alpha + \sin^3 \alpha} - \varrho (\cos \alpha + \sin \alpha)$$

$$= \varrho \left[\sqrt[3]{\cos^3 \alpha + \sin^3 \alpha} - (\cos \alpha + \sin \alpha)\right],$$

and we see that $\Delta z - (dx + dy)$ is of the same order as ϱ for the α for which the factor in square brackets differs from zero. Hence $z = \sqrt[3]{x^3 + y^3}$ has no differential at $P_0(0, 0)$, in spite of the existence of derivatives with respect to any direction at this point. The reader may easily verify that the partial derivatives of the function are discontinuous at $P_0(0, 0)$.

As a matter of fact, if we require the *continuity* as well as existence of the partial derivatives at a point, the existence of the differential now follows, i.e. the function is *differentiable.*

THEOREM. **If the function $z = f(x, y)$ has continuous partial derivatives $f'_x(x, y)$ and $f'_y(x, y)$ at a point $P(x, y)$, it is differentiable at this point.**

This theorem provides a sufficient test for the differentiability of a function of two independent variables.

Proof. We rewrite the formula Δz with $f(x, y + \Delta y)$ added and subtracted on the right-hand side:

$$\Delta z = [f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y)] + \\ + [f(x, y + \Delta y) - f(x, y)].$$

The expression in the first bracket is the increment of $f(x, y)$ when x receives the increment Δx and the second argument $y + \Delta y$ remains constant. We regard this as the increment of a function of x only and apply Lagrange's formula (Sec. 65). We have:

$$f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y) = f'_x(x + \theta_1 \Delta x, y + \Delta y)\, \Delta x$$

where $0 < \theta_1 < 1$. Similarly, on applying Lagrange's formula to the expression in the second bracket as the increment of a function of y only, we obtain:

$$f(x, y + \Delta y) - f(x, y) = f'_y(x, y + \theta_2 \Delta y)\, \Delta y \text{ where } 0 < \theta_2 < 1.$$

Therefore

$$\Delta z = f'_x(x + \theta_1 \Delta x, y + \Delta y)\, \Delta x + f'_y(x, y + \theta_2 \Delta y)\, \Delta y.$$

But f'_x and f'_y are continuous functions by hypothesis, so that if we put

$$f'_x(x + \theta_1 \Delta x, y + \Delta y) = f'_x(x, y) + \varepsilon_1,$$

$$f'_y(x, y + \theta_2 \Delta y) = f'_y(x, y) + \varepsilon_2,$$

ε_1 and ε_2 will tend to zero along with Δx and Δy. Thus

$$\Delta z = [f'_x(x, y) + \varepsilon_1]\, \Delta x + [f'_y(x, y) + \varepsilon_2]\, \Delta y,$$

or

$$\Delta z = f'_x(x, y)\, \Delta x + f'_y(x, y)\, \Delta y + \alpha, \qquad (*)$$

where $\alpha = \varepsilon_1 \Delta x + \varepsilon_2 \Delta y$.

We notice that α is an infinitesimal of higher order than the distance ϱ between points $P(x, y)$ and $P_1(x + \Delta x, y + \Delta y)$. For, since $|\Delta x| \leqslant \sqrt{\Delta x^2 + \Delta y^2}$ and $|\Delta y| \leqslant \sqrt{\Delta x^2 + \Delta y^2}$, we have

$$|\alpha| = |\varepsilon_1 \Delta x + \varepsilon_2 \Delta y| \leqslant |\varepsilon_1|\,|\Delta x| + |\varepsilon_2|\,|\Delta y| \leqslant$$

$$\leqslant |\varepsilon_1| \sqrt{\Delta x^2 + \Delta y^2} + |\varepsilon_2| \sqrt{\Delta x^2 + \Delta y^2} = (|\varepsilon_1| + |\varepsilon_2|)\varrho$$

i.e.
$$\frac{|\alpha|}{\varrho} \leqslant |\varepsilon_1| + |\varepsilon_2|.$$

If $\varrho \to 0$, then $\Delta x \to 0$ and $\Delta y \to 0$, which implies that ε_1 and ε_2 tend to zero, and therefore all the more $\alpha/\varrho \to 0$. But this means that α is an infinitesimal of higher order than ϱ. Hence the sum of the first two terms on the right-hand side of equation (*), which is linear in Δx and Δy, in fact represents by definition the differential of the function at the point $P(x, y)$.

4. Rules for Differentiation

150. Differentiation of a Function of a Function.

I. Let $z = F(x, y)$ be given as a function of a function via the intermediate variables u and v:

$$z = f(u, v),$$

u and v being functions of x and y:

$$u = \varphi(x, y), \quad v = \psi(x, y).$$

Thus

$$z = F(x, y) = f[\varphi(x, y), \psi(x, y)].$$

We assume here that functions f, φ and ψ have continuous partial derivatives.

THEOREM. **The partial derivative of a function of a function is equal to the sum of the products of the partial derivatives of the given function with respect to the intermediate arguments (u and v) with the partial derivatives of these arguments (u and v) with respect to the corresponding independent variables (x or y):**

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x}, \quad \frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y}.$$

Proof. We give x and y increments Δx and Δy; these produce increments Δu and Δv in the intermediate arguments $u = \varphi(x, y)$

and $v = \psi(x, y)$, which in turn change the value of function z by an amount Δz. On regarding z as a function $f(u, v)$ of variables u and v, we can write:

$$\Delta z = f'_u \, \Delta u + f'_v \, \Delta v + \alpha_1,$$

where (see Sec. 149, formula (*)):

$$\alpha_1 = \varepsilon'_1 \Delta u + \varepsilon'_2 \Delta v,$$

and $\varepsilon'_1 \to 0$ and $\varepsilon'_2 \to 0$ as $\Delta u \to 0$ and $\Delta v \to 0$, i.e. as $\varrho = \sqrt{\Delta x^2 + \Delta y^2} \to 0$.

Similarly, we have for u and v as functions of x and y:

$$\Delta u = \varphi'_x \Delta x + \varphi'_y \Delta y + \alpha_2,$$

$$\Delta v = \psi'_x \Delta x + \psi'_y \Delta y + \alpha_3,$$

where

$$\alpha_2 = \varepsilon''_1 \Delta x + \varepsilon''_2 \Delta y, \quad \alpha_3 = \varepsilon'''_1 \Delta x + \varepsilon'''_2 \Delta y,$$

and ε''_1, ε''_2, ε'''_1, ε'''_2 tend to zero as $\varrho = \sqrt{\Delta x^2 + \Delta y^2} \to 0$.

On substituting these expressions for Δu and Δv in the expression for Δz, we get

$$\Delta z = [(f'_u \varphi'_x + f'_v \psi'_x) \Delta x + (f'_u \varphi'_y + f'_v \psi'_y) \Delta y] + \alpha, \qquad (*)$$

where

$$\alpha = \varepsilon_1 \Delta x + \varepsilon_2 \Delta y,$$

$$\varepsilon_1 = \varepsilon''_1 f'_u + \varepsilon'''_1 f'_v + \varepsilon'_1 \varphi'_x + \varepsilon'_2 \psi'_x + \varepsilon'_1 \varepsilon''_1 + \varepsilon'_2 \varepsilon'''_1,$$

and

$$\varepsilon_2 = \varepsilon''_2 f'_u + \varepsilon'''_2 f'_v + \varepsilon'_1 \varphi'_y + \varepsilon'_2 \psi'_y + \varepsilon'_1 \varepsilon''_2 + \varepsilon'_2 \varepsilon'''_2.$$

These expressions show that $\varepsilon_1 \to 0$ and $\varepsilon_2 \to 0$ as $\varrho \to 0$ (i.e. as $\Delta x \to 0$, $\Delta y \to 0$).

Hence (see Sec. 149) the expression in square brackets on the right-hand side of equation (*) is the differential dz:

$$dz = (f'_u \varphi'_x + f'_v \psi'_x) dx + (f'_u \varphi'_y + f'_v \psi'_y) dy. \qquad (**)$$

It follows from this that

$$\frac{\partial z}{\partial x} = f'_u \varphi'_x + f'_v \psi'_x, \quad \frac{\partial z}{\partial y} = f'_u \varphi'_y + f'_v \psi'_y.$$

This is what we had to prove.

It will be seen that the rule for differentiation of a function of a function of two variables is similar to the rule for differentiation of a function of a function of one variable.

THEOREM. **The differential of a function $z = f(u, v)$ retains the same form independently of whether its arguments u and v are independent variables or functions of the independent variables.**

Proof. We regroup the terms in expression (**) for the differential:

$$dz = \frac{\partial z}{\partial u}\left(\frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\right) + \frac{\partial z}{\partial v}\left(\frac{\partial v}{\partial x}\,\partial x + \frac{\partial v}{\partial y}\,dy\right).$$

Thus

$$dz = \frac{\partial z}{\partial u}\,du + \frac{\partial z}{\partial v}\,dv.$$

This is what we wished to prove.

Hence the property of invariance of the form of the first differential (see Sec. 52) also holds for functions of two independent variables.

Example 1. $z = e^{xy}\sin(x + y)$. Putting $xy = u$, $x + y = v$, so that $z = e^u \sin v$, we find that

$$z'_x = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x} = e^u \sin v \cdot y + e^u \cos v \cdot 1$$

$$= e^{xy}\,[y \sin(x + y) + \cos(x + y)],$$

$$z'_y = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y} = e^u \sin v \cdot x + e^u \cos v \cdot 1$$

$$= e^{xy}\,[x \sin(x + y) + \cos(x + y)].$$

Example 2. A function $z = f(x, y)$ can be regarded as a function of a function of the polar co-ordinates r and φ of the point $P(x, y)$. We have $x = r\cos\varphi$, $y = r\sin\varphi$, so that

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = f'_x \cos\varphi + f'_y \sin\varphi.$$

This is the formula for the derivative of $z = f(x, y)$ with respect to the direction of the radius vector r of the point $P(x, y)$ already obtained in Sec. 148. Furthermore,

$$\frac{\partial z}{\partial \varphi} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial \varphi} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial \varphi} = -f'_x r \sin\varphi + f'_y r \cos\varphi.$$

II. The above rule for differentiation of a function of a function still holds for functions of any number of independent variables and for any number of intermediate arguments.

Let z be given directly as a function of arguments $u, v, \ldots, w$, which are functions of the independent variables $x, y, \ldots, t$. Then

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x} + \cdots + \frac{\partial z}{\partial w}\frac{\partial w}{\partial x},$$

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y} + \cdots + \frac{\partial z}{\partial w}\frac{\partial w}{\partial y},$$

. . . . . .

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial t} + \cdots + \frac{\partial z}{\partial w}\frac{\partial w}{\partial t}.$$

The expression for the differential of a function also remains the same independently of whether or not the initial arguments of the function are independent variables. In fact:

$$dz = \frac{\partial z}{\partial u}\,du + \frac{\partial z}{dv}\,\partial v + \cdots + \frac{\partial z}{\partial w}\,dw.$$

In the particular case when all the arguments $u, v, \ldots, w$ are functions of one independent variable x, we are in fact concerned with a function of x only. The (ordinary) derivative of the function—which is termed in this case the total derivative—now has the form

$$\frac{dz}{dx} = \frac{\partial z}{\partial u}\frac{du}{dx} + \frac{\partial z}{\partial v}\frac{dv}{dx} + \cdots + \frac{\partial z}{\partial w}\frac{dw}{dx}. \qquad (***)$$

Example. Let

$$y = (\tan x)^{\sin x}.$$

We put $\tan x = u$, $\sin x = v$. Then $y = u^v$. We have:

$$y' = \frac{\partial y}{\partial u}u' + \frac{\partial y}{\partial v}v' = vu^{v-1}\frac{1}{\cos^2 x} + u^v \ln u \cos x$$

$$= u^v\left(\frac{v}{u}\frac{1}{\cos^2 x} + \ln u \cos x\right)$$

$$= (\tan x)^{\sin x}\left(\frac{1}{\cos x} + \cos x \ln \tan x\right).$$

A special method—"logarithmic differentiation"—was recommended earlier for finding the derivative of such a function.

If we put $u = x$ in (***), we get

$$\frac{dz}{dx} = \frac{\partial z}{\partial x} + \frac{\partial z}{\partial v}\frac{dv}{dx} + \cdots + \frac{\partial z}{\partial w}\frac{dw}{dx}.$$

We must draw the reader's attention to the difference between the two derivatives with respect to x contained in this expression. Whilst dz/dx is the "total derivative", i.e. the ordinary derivative with respect to x of z as a function of x only, $\partial z/\partial x$ is the partial derivative of z with respect to the argument x appearing directly in the expresssion for the function, i.e. on the assumption that the remaining arguments remain constant during differentiation in spite of the fact that they depend on x.

For example, if

$$z = x^3 e^{u^2}, \quad \text{where} \quad u = \varphi(x),$$

then

$$\frac{\partial z}{\partial x} = 3x^2 e^{u^2}, \quad \text{whilst} \quad \frac{dz}{dx} = 3x^2 e^{u^2} + x^3 e^{u^2} 2u\, \varphi'(x).$$

III. Simple rules analogous to those for the case of one variable (Sec. 52) can be used when finding the differential of a function of several independent variables. Let $u, v, \ldots, w$ be functions of any number of independent variables. The rules expressed by the following formulae now hold:

(1) $d(u + v + \cdots + w) = du + dv + \cdots + dw$;

(2) $d(uv) = u\,dv + v\,du$, in particular, $d(cu) = c\,du$;

(3) $d\left(\frac{u}{v}\right) = \frac{v\,du - u\,dv}{v^2}$;

(4) $df[\varphi(u, v, \ldots, w)] = f'[\varphi(u, v, \ldots, w)]\, d\varphi(u, v, \ldots, w)$.

These formulae follow at once from the property of invariance of the form of the first differential.

We shall prove the second formula for illustration. Since the form of the differential does not depend on the nature of the arguments, we assume that these are independent variables. Then

$$d(uv) = \frac{\partial}{\partial v}(uv)\, dv + \frac{\partial}{\partial u}(uv)\, du = u\,dv + v\,du.$$

This is what we wanted to show.

The remaining rules are proved in a similar manner. Sometimes working can be greatly simplified by using these rules.

Example. Let us find $d \arctan y/x$. We have by the fourth rule:

$$d \arctan \frac{y}{x} = \left(\arctan \frac{y}{x}\right)' d\left(\frac{y}{x}\right) = \frac{1}{1 + \frac{y^2}{x^2}}\, d\left(\frac{y}{x}\right).$$

We use the third rule and finally obtain:

$$d \arctan \frac{y}{x} = \frac{1}{1 + \frac{y^2}{x^2}} \cdot \frac{x\,dy - y\,dx}{x^2} = -\frac{y}{x^2 + y^2}\,dx + \\ + \frac{x}{x^2 + y^2}\,dy.$$

It follows at once from this that

$$\frac{\partial}{\partial x} \arctan \frac{y}{x} = -\frac{y}{x^2 + y^2}, \quad \frac{\partial}{\partial y} \arctan \frac{y}{x} = \frac{x}{x^2 + y^2}.$$

151. Implicit Functions and their Differentiation.

I. EXPLICIT AND IMPLICIT FUNCTIONS. *We describe z as an explicit* function *of two independent variables x* and *y if it is given by an equation between x, y and z which is solved for z.* An insoluble equation between three variables can, however, define one of them as a function of the other two. In fact, the general form of equation between x, y and z is

$$f(x, y, z) = 0.$$

On substituting any given values $x = x_0$, $y = y_0$ for x and y, we obtain an equation in z:

$$f(x_0, y_0, z) = 0,$$

from which the value (or values) of z can be found corresponding to $x = x_0$ and $y = y_0$. Hence z is given as a definite function of x and y by the equation $f(x, y, z) = 0$. Such a function is said to be implicit.

Definition. A function defined by an equation between x, y and z and not solved with respect to z is said to be an implicit function *z of the two independent variables x and y.*

For instance, each of the equations

$$x^2 + y^2 + z^2 - R^2 = 0,$$
$$z^5 - 3xz^2 + 2y^2 - z + 1 = 0$$

defines z as an implicit function of x and y*.

* It must not be thought that equating any function $f(x, y, z)$ to zero gives an equation which necessarily defines a function of two variables. Examples are easily adduced of equations between x, y and z having no solutions. For instance, $x^2 + y^2 + z^2 + 1 = 0$ cannot be satisfied by real values of x, y and z (the sum of positive numbers is always greater than zero!), i.e. this equation does not define a function.

A theorem may be mentioned which indicates the conditions in which the equation $f(x, y, z) = 0$ defines z as a continuous and differentiable function of x and y:

THEOREM ON THE EXISTENCE OF AN IMPLICIT FUNCTION*. *If $f(x, y, z)$ is defined and continuous in a neighbourhood of the point $M_0(x_0, y_0, z_0)$, where $f(x_0, y_0, z_0) = 0$, whilst its partial derivatives $\partial f/\partial x$, $\partial f/\partial y$, $\partial f/\partial z$ exist and are continuous in the neighbourhood and $\partial f/\partial z$ does not vanish at the point M_0, the equation*

$$f(x, y, z) = 0$$

defines z as a single-valued and continuous function of x and y in some neighbourhood of the point $P_0(x_0, y_0)$:

$$z = \varphi(x, y),$$

taking the value z_0 at P_0 and having continuous partial derivatives $\partial z/\partial x$, $\partial z/\partial y$.

On solving the equation $f(x, y, z) = 0$ for z (if this is possible) we arrive at the explicit expression for z:

$$z = \varphi(x, y).$$

Substitution of this in the function $f(x, y, z)$ causes it to vanish identically (i.e. for any of the values of x and y):

$$f[x, y, \varphi(x, y)] = 0.$$

It is not possible to express z as an elementary function of x and y from every equation between x, y and z. This fact does not play an important role in analysis, however. If a single-valued function $z = \varphi(x, y)$ *exists*, defined by the given equation $f(x, y, z) = 0$ (and we shall in fact only be concerned with such cases), all the operations of mathematical analysis can be applied to it with equal success independently of whether or not it is expressed with the aid of an elementary function.

Definition. The function u defined by the equation

$$f(x, y, z, \ldots, t, u) = 0,$$

where $f(x, y, z, \ldots, t, u)$ is a function of the $n+1$ arguments x, y, z, $\ldots$, t, u is termed an implicit function *of the n independent variables x, y, z, $\ldots$, t.*

* For the proof see e.g. G.M. FIKHTENGOL'TS, *Course of Differential and Integral Calculus* (*Kurs differentsial'nogo i integral'nogo ischisleniya*), Vol. 1, p. 508 *et seq.*; Gost., 1947, R. COURANT, *Course of Differential and Integral Calculus*. Part. II.

Everything that has been said regarding implicit functions of two independent variables can be carried over directly to an implicit function of any number of independent variables, as also to functions of a single variable.

In particular, the "existence theorem" holds. We shall state this for functions of one variable.

If $f(x, y)$ is defined and continuous in a neighbourhood of a point $M_0(x_0, y_0)$, where $f(x_0, y_0) = 0$, whilst its partial derivatives $\partial f/\partial x$, $\partial f/\partial y$ exist and are continuous in the neighbourhood and $\partial f/\partial y$ does not vanish at M_0, the equation

$$f(x, y) = 0$$

defines y in a neighbourhood of the point $P_0(x_0)$ as a single-valued and continuous function of x, $y = \varphi(x)$, taking the value y_0 at P_0 and having a continuous derivative.

The reader should formulate the existence theorem for an implicit function in the case of any number of arguments.

It will always be assumed in future that the conditions of the existence theorem are fulfilled.

II. Differentiation of implicit functions. Let the equation

$$f(x, y) = 0$$

define y as a single-valued and differentiable function $y = \varphi(x)$ of the independent variable x. If we substitute the function $\varphi(x)$ for y in the equation, we get the identity

$$f[x, \varphi(x)] = 0.$$

Thus the derivative with respect to x of the function $f(x, y)$, where $y = \varphi(x)$, must also vanish. We find on applying the rule for differentiation of a function of a function (Sec. 150):

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0,$$

whence

$$\frac{dy}{dx} = y' = -\frac{\dfrac{\partial f}{\partial x}}{\dfrac{\partial f}{\partial y}}.$$

This formula provides *the general expression for the derivative of an implicit function of one independent variable* (see Sec. 49).

Example 1. $f(x, y) \equiv \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0.$

We have $\partial f/\partial x = 2x/a^2$, $\partial f/\partial y = 2y/b^2$; hence

$$y' = -\frac{\frac{2x}{a^2}}{\frac{2y}{b^2}} = -\frac{b^2}{a^2}\frac{x}{y}.$$

Example 2. $f(x, y) \equiv xy - e^x + e^y = 0.$

Here $\partial f/\partial x = y - e^x$, $\partial f/\partial y = x + e^y$. Hence

$$y' = -\frac{y - e^x}{x + e^y}.$$

Now let the equation

$$f(x, y, z) = 0$$

define z as a single-valued and differentiable function $z = \varphi(x, y)$ of the independent variables x and y. If we replace z by the function $\varphi(x, y)$ in the equation, we get the identity

$$f[x, y, \varphi(x, y)] = 0.$$

Hence the partial derivatives with respect to x and y of the function $f(x, y, z)$, where $z = \varphi(x, y)$, must also vanish. We find by differentiation:

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 0 \quad \text{and} \quad \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial y} = 0,$$

whence

$$z'_x = \frac{\partial z}{\partial x} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial z}}, \quad z'_y = \frac{\partial z}{\partial y} = -\frac{\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial z}}.$$

These expressions give *the general formulae for the partial derivatives of an implicit function of two independent variables.*

Example 1. Let us find the partial derivatives of the function z given by the equation $f(x, y, z) \equiv x^2 + y^2 + z^2 - R^2 = 0$ (the equation of a sphere). We have:

$$\frac{\partial f}{\partial x} = 2x, \quad \frac{\partial f}{\partial y} = 2y, \quad \frac{\partial f}{\partial z} = 2z;$$

i. e.

$$z'_x = -\frac{x}{z}, \quad z'_y = -\frac{y}{z}.$$

In this example we can find the explicit form of the function and verify the results obtained.

Example 2. $f(x, y, z) \equiv e^{-xy} - 2z + e^z = 0$. We find:

$$\frac{\partial f}{\partial x} = -ye^{-xy}, \quad \frac{\partial f}{\partial y} = -xe^{-xy}, \quad \frac{\partial f}{\partial z} = -2 + e^z.$$

Therefore

$$z'_x = \frac{ye^{-xy}}{e^z - 2}, \quad z'_y = \frac{xe^{-xy}}{e^z - 2}.$$

The explicit form of z cannot be found in this case, and the values of the partial derivatives are obtained in terms of the independent variables x and y and the function z itself.

Example 3. The following general proposition used in thermodynamics may easily be proved:

If three variables x, y, z are interconnected by the relationship $f(x, y, z) = 0$, the existence theorem for an implicit function being valid for each of the variables, we have

$$\frac{\partial x}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial x} = -1.$$

This is proved simply by writing down the expressions for the partial derivatives concerned:

$$\frac{\partial x}{\partial y} = -\frac{f'_y(x, y, z)}{f'_x(x, y, z)}, \quad \frac{\partial y}{\partial z} = -\frac{f'_z(x, y, z)}{f'_y(x, y, z)}, \quad \frac{\partial z}{\partial x} = -\frac{f'_x(x, y, z)}{f'_z(x, y, z)},$$

whence the required relationship is at once seen to hold.

We deduced this relationship in Sec. 145 for the particular case when $(x, y, z) \equiv xy - Rz$.

In the general case when the equation

$$f(x, y, \ldots, u) = 0$$

defines u as a single-valued and differentiable function of x, y, ..., we find as above that

$$u'_x = -\frac{f'_x}{f'_u}, \quad u'_y = -\frac{f'_y}{f'_u}, \ldots$$

152. Functions Given in the Parametric Form and their Differentiation. I. FUNCTIONS GIVEN IN THE PARAMETRIC FORM. A function of several variables can also be written *parametrically*. We shall dwell only on functions of two independent variables. *Two* parameters are required to give such functions in the parametric form.

Let z be given as a function of a function of x and y via the intermediate variables u and v:

$$z = f(u, v),$$

where

$$u = F(x, y), \quad v = \Phi(x, y).$$

We obtain on solving (if possible) the latter system of two equations for x and y:

$$x = \varphi(u, v), \quad y = \psi(u, v),$$

where φ and ψ denote certain functions of u and v.

The specification of the function by a system of three equations

$$z = f(u, v), \quad u = F(x, y), \quad v = \Phi(x, y)$$

can be replaced by the equivalent system

$$x = \varphi(u, v), \quad y = \psi(u, v), \quad z = f(u, v).$$

We see that all three variables x, y, z, interconnected by a functional relationship, are expressed as functions of *two common parameters. The function of two independent variables thus defined is said to be given parametrically.*

As in the case of functions of one variable, the parametric form of a function of two independent variables often proves very convenient. This is due to the fact that it can eliminate the many-valuedness of the expressions connecting the three variables directly.

Example. The function z of two variables x and y defined by

$$x^2 + y^2 + z^2 - R^2 = 0$$

(the equation of the sphere of radius R with centre at the origin) is two-valued. This function may easily be shown to be expressible say by parameters φ and θ as follows:

$$x = R \sin\theta \cos\varphi, \quad y = R \sin\theta \sin\varphi, \quad z = R \cos\theta$$

$$0 \leqslant \varphi < 2\pi, \quad 0 \leqslant \theta \leqslant \pi.$$

This is in many cases more convenient than the original form since the functions of two variables here expressing variables x, y and z are single-valued.

The parameters φ and θ are interpreted geometrically as the spherical co-ordinates of a point of the sphere.

Generally, *the position of a point M in space is defined in a system of spherical co-ordinates* by the distance* ϱ *of M from the origin (the radius vector), the angle* φ *between the projection of the radius vector on the Oxy plane and the Ox axis (the longitude) and the angle* θ *between the radius vector and the Oz axis* ($\frac{1}{2}\pi - \theta$ *is the latitude)* (Fig. 16). Here, φ can vary from 0 to 2π, and θ from 0 to π (or from $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$).

The equation of a sphere with centre at the origin is obviously $\varrho = R = \text{const.}$ in the system of sperical co-ordinates.

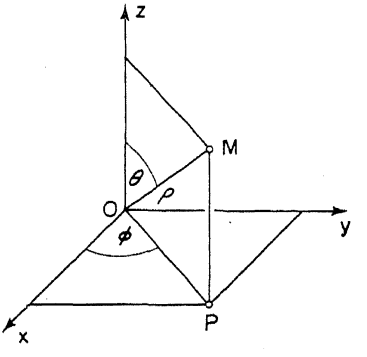

FIG. 16

II. DIFFERENTIATION OF FUNCTIONS GIVEN IN THE PARAMETRIC FORM.

Let the equations

$$x = \varphi(u, v), \quad y = \psi(u, v), \quad z = f(u, v)$$

define one of the variables x, y, z (say z) as a function of the other two (x and y), functions φ, ψ, f being differentiable. Let us find z'_x and z'_y. We differentiate the equation $z = f(u, v)$, bearing in mind that parameters u and v are functions of x and y given by the system of two equations:

$$x = \varphi(u, v), \quad y = \psi(u, v).$$

We have:

$$z'_x = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x}, \quad z'_y = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y}.$$

We find the derivatives $\partial u/\partial x$, $\partial v/\partial x$, $\partial u/\partial y$, $\partial v/\partial y$ from the systems of equations which are obtained after differentiation of the equations $x = \varphi(u, v)$ and $y = \psi(u, v)$ with respect to x and y. For example, differentiation with respect to x gives

$$1 = \frac{\partial x}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial x}{\partial v}\frac{\partial v}{\partial x},$$

$$0 = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial v}{\partial x}$$

(y is independent of x so that $\partial y/\partial x = 0$). This system gives us expressions for $\partial u/\partial x$ and $\partial v/\partial x$. A similar procedure is used for finding $\partial u/\partial y$ and $\partial v/\partial y$ (cf. Sec. 55, II).

Example. Let

$$x = R\sin\theta\cos\varphi, \quad y = R\sin\theta\sin\varphi, \quad z = R\cos\theta.$$

* Spherical co-ordinates are sometimes called polar co-ordinates in space.

We have:

$$z'_x = -R\sin\theta\,.\,\theta'_x, \quad z'_y = -R\sin\theta\,.\,\theta'_y.$$

Differentiation of the first two equations gives the two systems:

$$\left.\begin{aligned} 1 &= R\cos\theta\cos\varphi\,.\,\theta'_x - R\sin\theta\sin\varphi\,.\,\varphi'_x,\\ 0 &= R\cos\theta\sin\varphi\,.\,\theta'_x + R\sin\theta\cos\varphi\,.\,\varphi'_x;\end{aligned}\right\}$$

$$\left.\begin{aligned} 0 &= R\cos\theta\cos\varphi\,.\,\theta'_y - R\sin\theta\sin\varphi\,.\,\varphi'_y,\\ 1 &= R\cos\theta\sin\varphi\,.\,\theta'_y + R\sin\theta\cos\varphi\,.\,\varphi'_y.\end{aligned}\right\}$$

We find from the first system:

$$\theta'_x = \frac{\cos\varphi}{R\cos\theta},$$

and from the second:

$$\theta'_y = \frac{\sin\varphi}{R\cos\theta}.$$

Substitution in the expressions for z'_x and z'_y gives us:

$$z'_x = -\tan\theta\cos\varphi, \quad z'_y = -\tan\theta\sin\varphi.$$

These expressions for z'_x and z'_y are readily seen to coincide with those found earlier (Sec. 151).

The equation for the tangent plane to the sphere at the point $M_0(x_0, y_0, z_0)$ may be written as

$$z - z_0 = -\tan\theta_0\cos\varphi_0\,.\,(x - x_0) - \tan\theta_0\sin\varphi_0\,.\,(y - y_0),$$

where θ_0 and φ_0 are the values of θ and φ corresponding to the point M_0. Hence

$$\sin\theta_0\cos\varphi_0\,.\,(x - x_0) + \sin\theta_0\sin\varphi_0\,.\,(y - y_0) + \\ + \cos\theta_0\,.\,(z - z_0) = 0.$$

The coefficients of x, y, and z in this equation are readily seen to be the respective direction cosines of the radius vector of the point M_0; therefore, in accordance with the familiar fact of analytic geometry, the tangent plane to the sphere is perpendicular to the radius passing through the point of contact.

5. Repeated Differentiation

153. Derivatives of Higher Orders. Suppose that the function $z = f(x, y)$ has partial derivatives

$$\frac{\partial z}{\partial x} = f'_x(x, y), \quad \frac{\partial z}{\partial y} = f'_y(x, y),$$

which are continuous functions in some domain of the independent variables x and y. The partial derivatives of these functions (if they exist) are called the second partial derivatives or partial derivatives of the second order of the given function $f(x, y)$. Each first order derivative $(\partial z/\partial x, \partial z/\partial y)$ has two partial derivatives; we therefore obtain four partial derivatives of the second order, which are written as

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 z}{\partial x^2} = f''_{x^2} = z''_{x^2}, \quad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 x}{\partial x \partial y} = f''_{xy} = z''_{xy},$$

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial y \partial x} = f''_{yx} = z''_{yx}, \quad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial y^2} = f''_{y^2} = z''_{y^2}.$$

We refer to f''_{xy} and f''_{yx} as mixed derivatives; one is got by differentiating the function first with respect to x, then with respect to y, whilst the other is got by first differentiating with respect to y, then with respect to x.

Example. We have for the function $z = x^3y^2 - 3xy^3 - xy + 1$:

$$\frac{\partial z}{\partial x} = 3x^2y^2 - 3y^3 - y, \quad \frac{\partial z}{\partial y} = 2x^3y - 9xy^2 - x,$$

$$\frac{\partial^2 z}{\partial x^2} = 6xy^2, \quad \frac{\partial^2 z}{\partial y^2} = 2x^3 - 18xy,$$

$$\frac{\partial^2 z}{\partial x \partial y} = 6x^2y - 9y^2 - 1, \quad \frac{\partial^2 z}{\partial y \partial x} = 6x^2y - 9y^2 - 1.$$

It will be noticed that the mixed derivatives are identical here. This is not a chance occurrence.

THEOREM. **Given the continuity of the mixed second derivatives of a function $z = f(x, y)$ at a point $P(x, y)$, the derivatives must be equal at this point.***

* Continuity of the partial derivatives is an essential condition; the theorem may not hold if it is not fulfilled.

Proof. We consider the expression

$$A = f(x + \Delta x, y + \Delta y) - f(x + \Delta x, y) - f(x, y + \Delta y) + f(x, y).$$

We transform it by two different methods. We first put the terms into two groups whilst preserving their order:

$$A = [f(x + h, y + k) - f(x + h, y)] - [f(x, y + k) - f(x, y)],$$

where we have written for brevity $\Delta x = h$, $\Delta y = k$. Using the notation

$$f(x, y + k) - f(x, y) = \varphi(x)$$

(y is not indicated as an argument of the function φ since we are not at present interested in its variation), the expression in the first square bracket is easily seen to be $\varphi(x + h)$:

$$f(x + h, y + k) - f(x + h, y) = \varphi(x + h).$$

We thus obtain by using Lagrange's formula:

$$A = \varphi(x + h) - \varphi(x) = h\varphi'(x + \theta h),$$

where $0 < \theta < 1$. But

$$\varphi'(x) = f'_x(x, y + k) - f'_x(x, y).$$

We apply Lagrange's formula to this difference, bearing in mind that the second argument y varies here. We get

$$\varphi'(x) = kf''_{xy}(x, y + \theta_1 k), \quad 0 < \theta_1 < 1,$$

whence

$$\varphi'(x + \theta h) = kf''_{xy}(x + \theta h, \quad y + \theta_1 k),$$

and finally

$$A = hkf''_{xy}(x + \theta h, y + \theta_1 k). \qquad (*)$$

We now put the terms of the original expression for A into two different groups:

$$A = [f(x + h, y + k) - f(x, y + k)] - [f(x + h, y) - f(x, y)].$$

We write

$$f(x + h, y) - f(x, y) = \psi(y).$$

Now, in precisely the same way as above,

$$A = \psi(y + k) - \psi(y) = k\psi'(y + \theta' k), \quad 0 < \theta' < 1.$$

But

$$\psi'(y) = f'_y(x + h, y) - f'_y(x, y) = hf''_{yx}(x + \theta'_1 h, y), \quad 0 < \theta'_1 < 1$$

so that

$$\psi'(y + \theta' k) = hf''_{yx}(x + \theta'_1 h, y + \theta' k),$$

which finally gives

$$A = khf''_{yx}(x + \theta'_1 h, y + \theta' k). \qquad (**)$$

We equate expressions (*) and (**) obtained for A:

$$hkf''_{xy}(x + \theta h, y + \theta_1 k) = khf''_{yx}(x + \theta'_1 h, y + \theta' k).$$

Thus

$$f''_{xy}(x + \theta h, y + \theta_1 k) = f''_{yx}(x + \theta'_1 h, y + \theta' k).$$

We let h and k tend to zero; we have, in view of the assumed continuity of the second derivatives at the point $P(x, y)$:

$$\lim_{\substack{h \to 0 \\ k \to 0}} f''_{xy}(x + \theta h, y + \theta_1 k) = f''_{xy}(x, y),$$

$$\lim_{\substack{h \to 0 \\ k \to 0}} f''_{yx}(x + \theta'_1 h, y + \theta' k) = f''_{yx}(x, y),$$

and we arrive at the required equality:

$$f''_{xy}(x, y) = f''_{yx}(x, y).$$

Thus, given the conditions mentioned, a function of two variables $z = f(x, y)$ has three and not four second order partial derivatives:

$$\frac{\partial^2 z}{\partial x^2}, \quad \frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}, \quad \frac{\partial^2 z}{\partial y^2}.$$

The partial derivatives of the second order partial derivatives are termed third order or third partial derivatives.

Definition. ***A partial derivative of an $(n-1)$-th order partial derivative* is termed an n-th order or n-th partial derivative.**

The n-th order partial derivative of a function $z = f(x, y)$, taken k times with respect to x and $(n - k)$ times with respect to y, can be written in accordance with the order in which the differentiation is carried out:

$$\frac{\partial^n z}{\partial x^k \partial y^{n-k}}, \quad \frac{\partial^n z}{\partial y^{n-k} \partial x^k} \quad \text{etc.}$$

$$\left(\text{or} \quad f^{(n)}_{x^k y^{n-k}}(x, y), \quad f^{(n)}_{y^{n-k} x^k}(x, y) \quad \text{etc.}\right).$$

The theorem on the equality of the mixed second derivatives enables us to prove a general proposition:

The result of repeated differentiation of a function of two independent variables does not depend on the order of the differentiation (the partial derivatives in question are assumed to be *continuous*).

Let us show, for instance, that

$$\frac{\partial^3 z}{\partial x\,\partial y^2} = \frac{\partial^3 z}{\partial y\,\partial x\,\partial y}.$$

We find on using the theorem on the equality of the mixed second derivatives:

$$\frac{\partial^3 z}{\partial x\,\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial^2 z}{\partial x\,\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial^2 z}{\partial y\,\partial x}\right) = \frac{\partial^3 z}{\partial y\,\partial x\,\partial y}.$$

This general proposition can be proved similarly in all other possible cases.

Let an n-th order partial derivative be obtained for a function $z = f(x, y)$ by differentiating altogether k times with respect to x and $(n - k)$ times with respect to y; if the n-th order derivatives are continuous, the present derivative can be written as $\partial^n z/\partial x^k\,\partial y^{n-k}$ (or as $\partial^n z/\partial y^{n-k}\partial x^k$) independently of the order in which the differentiations are carried out.

A function $z = f(x, y)$ thus has in fact $(n + 1)$ partial derivatives of the n-th order, which can be denoted by

$$\frac{\partial^n z}{\partial x^n},\ \frac{\partial^n z}{\partial x^{n-1}\,\partial y},\ \frac{\partial^n z}{\partial x^{n-2}\,\partial y^2},\ \ldots,\ \frac{\partial^n z}{\partial x^2\,\partial y^{n-2}},\ \frac{\partial^n z}{\partial x\,\partial y^{n-1}},\ \frac{\partial^n z}{\partial y^n}.$$

The elementary functions of two independent variables generally speaking (i.e. except for individual points and individual curves) have partial derivatives of any order in their domain of definition.

Higher order partial derivatives may be defined similarly for functions of any number of independent variables. The theorem regarding the independence of the result on the order of differentiation still holds. For instance, if $u = f(x, y, z)$, then

$$\frac{\partial^3 u}{\partial x\,\partial y\,\partial z} = \frac{\partial^3 u}{\partial x\,\partial z\,\partial y} = \frac{\partial^3 u}{\partial y\,\partial x\,\partial z} = \frac{\partial^3 u}{\partial y\,\partial z\,\partial x}$$
$$= \frac{\partial^3 u}{\partial z\,\partial x\,\partial y} = \frac{\partial^3 u}{\partial z\,\partial y\,\partial x}.$$

Repeated differentiation of a function of several variables is carried out in practice by finding successively one derivative after another by the familiar rules of differentiation.

Example. Let us find the second partial derivatives $\partial^2 u/\partial x^2$, $\partial^2 u/\partial y^2$, $\partial^2 u/\partial z^2$ of the "inverse radius vector" of a point in space:

$$u = f(x, y, z) = \frac{1}{r} = \frac{1}{\sqrt{x^2 + y^2 + z^2}}.$$

We have:

$$\frac{\partial u}{\partial x} = \frac{\partial\left(\frac{1}{r}\right)}{\partial x} = \frac{\partial\left(\frac{1}{r}\right)}{\partial r}\frac{\partial r}{\partial x} = -\frac{1}{r^2}\frac{x}{r} = -\frac{x}{r^3},$$

$$\frac{\partial u}{\partial y} = -\frac{y}{r^3}, \quad \frac{\partial u}{\partial z} = -\frac{z}{r^3},$$

and further,

$$\frac{\partial^2 u}{\partial x^2} = -\frac{r^3 - 3r^2\frac{\partial r}{\partial x}x}{r^6} = -\frac{r^3 - 3rx^2}{r^6} = -\frac{r^2 - 3x^2}{r^5},$$

$$\frac{\partial^2 u}{\partial y^2} = -\frac{r^2 - 3y^2}{r^5}, \quad \frac{\partial^2 u}{\partial z^2} = -\frac{r^2 - 3z^2}{r^5}.$$

It will be noticed that the "inverse radius vector" satisfies the differential equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0,$$

which is known as *Laplace's equation.*

An interesting point is that the function $u = 1/r$ does not satisfy the analogous equation for a function of two variables in a plane $\left(r = \sqrt{x^2 + y^2}\right)$:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

We suggest that the reader show that this equation is satisfied e.g. by the function

$$u = \ln\frac{1}{r} = \ln\frac{1}{\sqrt{x^2 + y^2}}.$$

154. Differentials of Higher Orders. The differential of the function $z = f(x, y)$:

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$$

is a function of the independent variables x and y and of their differentials dx and dy. The differentials dx and dy are regarded as quantities independent of x and y.

The differential $d(dz)$ is known as the second order or second differential and is written as d^2z. The differential $d(d^2z)$ is known as the third order or third differential and so on. The following general definition may be given*.

* We shall not dwell on the partial differentials of higher orders.

Definition. **The *n*-th order or *n*-th *differential* $d^n z$ of a function $z = f(x, y)$ is defined as the differential of the $(n-1)$-th differential, regarded as a function of the independent variables x and y.**

We shall find an expression for $d^n z$ in terms of the partial derivatives of function z and the differentials of the independent variables. We find for the second order differential:

$$d(dz) = \frac{\partial}{\partial x}(dz)\,dx + \frac{\partial}{\partial y}(dz)\,dy$$

$$= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy\right)dx + \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy\right)dy,$$

and since dx and dy can be reckoned constant with respect to x and y by virtue of what has been said, we get:

$$d(dz) = \left(\frac{\partial^2 z}{\partial x^2}dx + \frac{\partial^2 z}{\partial y\,\partial x}dy\right)dx + \left(\frac{\partial^2 z}{\partial x\,\partial y}dx + \frac{\partial^2 z}{\partial y^2}dy\right)dy,$$

or, assuming the continuity of the mixed derivatives:

$$d(dz) = \frac{\partial^2 z}{\partial x^2}dx^2 + 2\frac{\partial^2 z}{\partial x\,\partial y}dx\,dy + \frac{\partial^2 z}{\partial y^2}dy^2.$$

We have for the third differential:

$$d(d^2 z) = \frac{\partial}{\partial x}(d^2 z)\,dx + \frac{\partial}{\partial y}(d^2 z)\,dy;$$

we find on carrying out the differentiations:

$$d^3 z = \frac{\partial^3 z}{\partial x^3}dx^3 + 3\frac{\partial^3 z}{\partial x^2\,\partial y}dx^2\,dy + 3\frac{\partial^3 z}{\partial x\,\partial y^2}dx\,dy^2 + \frac{\partial^3 z}{\partial y^3}dy^3.$$

It will be seen that the second and third differentials consist of $\partial z/\partial x \,.\, dx$ and $\partial z/\partial y \,.\, dy$ regarded as terms of a polynomial expressing $\partial z/\partial x \,.\, dx + \partial z/\partial y \,.\, dy$ to the second and third powers respectively.

It may easily be verified that in general, on passing from n to $n+1$:

$$d^n z = \frac{\partial^n z}{\partial x^n}dx^n + C_n^1\frac{\partial^n z}{\partial x^{n-1}\,\partial y}dx^{n-1}\,dy +$$

$$+ C_n^2\frac{\partial^n z}{\partial x^{n-2}\,\partial y^2}dx^{n-2}\,dy^2 + \cdots$$

$$\cdots + C_n^{n-1}\frac{\partial^n z}{\partial x\,\partial y^{n-1}}dx\,dy^{n-1} + \frac{\partial^n z}{\partial y^n}dy^n,$$

where C_n^1, C_n^2, ... are the binomial coefficients (the number of combinations of n elements 1 at a time, 2 at a time, ...). In view of this the n-th differential can be written symbolically as

$$d^n z = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^n z .$$

On expanding this "binomial" in accordance with the ordinary Newton formula and performing the operations of the symbols $\partial/\partial x$, $\partial/\partial y$, regarded as fractions, we obtain the expression for the differential if we write z behind the symbol ∂^k.

The n-th differential is a homogeneous polynomial of degree n in dx and dy (i.e. each term is of the n-th order in dx and dy).

If the n-th differential $d^n z$ does not vanish, it is an infinitesimal of order n compared with $\varrho = \sqrt{dx^2 + dy^2}$.

The successive differentials (not equal to zero) of the function $z = f(x, y)$:

$$dz, d^2z, d^3z, \ldots, d^n z, \ldots$$

are thus arranged in order of increasing smallness.

We observe finally that, if the initial arguments x and y of a function z are not *independent* variables, only the first differential retains the same form as would be got if x and y were independent (Sec. 150); all the higher order differentials take a new form. In fact, if x and y are assumed to be functions of two other independent variables, we cannot look on the differentials dx and dy as independent of these variables, so that they must be regarded as variables subject to differentiation. For instance,

$$d^2z = d\left(\frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy\right)$$

$$= d\left(\frac{\partial z}{\partial x}\right)dx + \frac{\partial z}{\partial x}d\,(dx) + d\left(\frac{\partial z}{\partial y}\right)dy + \frac{\partial z}{\partial y}d(dy)$$

$$= \frac{\partial^2 z}{\partial x^2}dx^2 + \frac{\partial^2 z}{\partial x\,\partial y}dx\,dy + \frac{\partial z}{\partial x}d^2x + \frac{\partial^2 z}{\partial y\,\partial x}dy\,dx +$$

$$+ \frac{\partial^2 z}{\partial y^2}dy^2 + \frac{\partial z}{\partial y}d^2y = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^2 z + \frac{\partial z}{\partial x}d^2x + \frac{\partial z}{\partial y}d^2y .$$

It will be seen that two additional terms appear, which vanish if x and y are independent variables, since then $d^2x = 0$ and $d^2y = 0$.

CHAPTER XI

APPLICATIONS OF THE DIFFERENTIAL CALCULUS

1. Taylor's Formula. The Extrema of a Function of Several Independent Variables

155. Taylor's Formula and Series for Functions of Several Variables.

I. We shall prove Taylor's theorem for functions of several independent variables. The case of two independent variables will be treated in detail.

TAYLOR'S THEOREM. **If a function $z = f(x, y)$ is continuous together with its partial derivatives up to and including the $(n+1)$-th order in a neighbourhood of a point $P_0(x_0, y_0)$, and $P(x, y)$ is any point of this neighbourhood, the formula holds:**

$$f(x, y) = f(x_0\, y_0) + \left[\frac{\partial}{\partial x}(x - x_0) + \frac{\partial}{\partial y}(y - y_0)\right] f_{\substack{x = x_0 \\ y = y_0}} +$$

$$+ \frac{1}{2!}\left[\frac{\partial}{\partial x}(x - x_0) + \frac{\partial}{\partial y}(y - y_0)\right]^2 f_{\substack{x = x_0 \\ y = y_0}} + \cdots$$

$$\cdots + \frac{1}{n!}\left[\frac{\partial}{\partial x}(x - x_0) + \frac{\partial}{\partial y}(y - y_0)\right]^n f_{\substack{x = x_0 \\ y = y_0}} +$$

$$+ \frac{1}{(n+1)!}\left[\frac{\partial}{\partial x}(x - x_0) + \frac{\partial}{\partial y}(y - y_0)\right]^{n+1} f_{\substack{x = \xi \\ y = \eta}}, \qquad (*)$$

where the point $P_c(\xi, \eta)$ lies on the straight line joining points P_0 and P.

Expression (*) is known as *Taylor's formula* of the n-th order for the function $f(x, y)$ of two independent variables at the point $P_0(x_0, y_0)$.

Proof. We take a point $P_1(x_1, y_1)$ in the neighbourhood of P_0. We put $x_1 - x_0 = \Delta x$, $y_1 - y_0 = \Delta y$ and consider the function

$$\varphi(t) = f(x_0 + t\,\Delta x,\ y_0 + t\,\Delta y)$$

of the parameter t. With fixed P_0 and P_1 this is a function of the single independent variable t. As t varies from zero to unity the point P moves along the straight line $x = x_0 + t(x_1 - x_0)$, $y = y_0 + t(y_1 - y_0)$, i.e. along the straight line joining P_0 and P_1. We have $\varphi(0) = f(x_0, y_0)$, $\varphi(1) = f(x_1, y_1)$. By hypothesis $\varphi(t)$ has continuous derivatives up to order $(n+1)$ at $t = 0$. We can therefore apply Taylor's formula to $\varphi(t)$ (see Sec. 78) at the point $t = 0$:

$$\varphi(t) = \varphi(0) + \varphi'(0)\, t + \frac{\varphi''(0)}{2!} t^2 + \cdots + \frac{\varphi^{(n)}(0)}{n!} t^n + \\ + \frac{\varphi^{(n+1)}(\theta)}{(n+1)!} t^{n+1},$$

where θ is a number lying between zero and t.

Let us now find the derivatives of the function

$$\varphi(t) = f(x, y),$$

where $x = x_0 + t\Delta x$, $y = y_0 + t\Delta y$. We obtain by the rule for differentiation of a function of a function:

$$\varphi'(t) = f'_x(x, y)\frac{dx}{dt} + f'_y(x, y)\frac{dy}{dt} = f'_x(x, y)\Delta x + f'_y(x, y)\,\Delta y;$$

we have for the second derivative:

$$\varphi''(t) = f''_{x^2}(x, y)\,\Delta x^2 + 2f''_{xy}(x, y)\,\Delta x\,\Delta y + f''_{y^2}(x, y)\,\Delta y^2;$$

and we obtain in general on passing from k to $k+1$:

$$\varphi^{(k)}(t) = f^{(k)}_{x^k}(x, y)\,\Delta x^k + C^1_k f^{(k)}_{x^{k-1}y}(x, y)\,\Delta x^{k-1}\,\Delta y + \\ + C^2_k f^{(k)}_{x^{k-2}y^2}(x, y)\,\Delta x^{k-2}\Delta y^2 + \cdots + C^{k-1}_k f^{(k)}_{xy^{k-1}}(x, y)\,\Delta x\,\Delta y^{k-1} + \\ + f^{(k)}_{y^k}(x, y)\,\Delta y^k.$$

On setting $t = 0$ in this expression, we arrive at the required expression for $\varphi^{(k)}(0)$. Since $x = x_0$ and $y = y_0$ for $t = 0$, $\varphi^{(k)}(0)$ is obtained from $\varphi^{(k)}(t)$ by substituting x_0 for x and y_0 for y in the partial derivatives. We can write symbolically (Sec. 154):

$$\varphi^{(k)}(0) = \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^k f_{\substack{x = x_0 \\ y = y_0}}, \quad k = 0, 1, \ldots, n.$$

The subscripts show that the partial derivatives are taken for $x = x_0$, $y = y_0$.

As regards the derivative $\varphi^{(n+1)}(\theta)$, its value is got from the expression for $\varphi^{(n+1)}(t)$ by replacing x and y by $\xi = x_0 + \theta \Delta x$, $\eta = y_0 + \theta \Delta y$ respectively. The point $P(\xi, \eta)$ lies on the straight line joining points $P_0(x_0, y_0)$ and $P_1(x_1, y_1)$.

Therefore,

$$\varphi(t) = f(x_0, y_0) + \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right) f_{\substack{x=x_0\\y=y_0}} t + \\ + \frac{1}{2!}\left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^2 f_{\substack{x=x_0\\y=y_0}} t^2 + \cdots \\ \cdots + \frac{1}{n!}\left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^n f_{\substack{x=x_0\\y=y_0}} t^n + \\ + \frac{1}{(n+1)!}\left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^{n+1} f_{\substack{x=\xi\\y=\eta}} t^{n+1}.$$

We obtain by putting $t = 1$ here:

$$\varphi(1) = f(x_1, y_1) = f(x_0, y_0) + \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right) f_{\substack{x=x_0\\y=y_0}} + \\ + \frac{1}{2!}\left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^2 f_{\substack{x=x_0\\y=y_0}} + \cdots \\ \cdots + \frac{1}{n!}\left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^n f_{\substack{x=x_0\\y=y_0}} + \\ + \frac{1}{(n+1)!}\left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^{n+1} f_{\substack{x=\xi\\y=\eta}}.$$

On omitting the subscripts from the co-ordinates of the point at which the value of the function is found and replacing Δx and Δy by $x - x_0$ and $y - y_0$, we obtain formula(*), which is what we wanted to prove.

We now note that the symbolic binomials on the right-hand side of Taylor's formula (*) are in fact the successive differentials of the function $f(x, y)$. Hence, on first taking $f(x_0, y_0)$ over to the left-hand side, we can rewrite (*) as

$$\Delta f(P_0) = df(P_0) + \frac{1}{2!} d^2 f(P_0) + \cdots + \frac{1}{n!} d^n f(P_0) + \\ + \frac{1}{(n+1)!} d^{n+1} f(P_c), \quad (**)$$

where $P_c = P_c(x_0 + \theta \Delta x, y_0 + \theta \Delta y)$ denotes a point lying on the straight line joining the initial point $P_0(x_0, y_0)$ and the end point $P(x_0 + \Delta x, y_0 + \Delta y)$.

This form of Taylor's formula for a function of two independent variables is precisely the same as the formula for a function of one variable (see Sec. 78). In general:

Equation (**) is *Taylor's formula* for a function of any number of variables.

We get from (**) with $n = 0$:

$$\Delta f(P_0) = df(P_c),$$

i.e. *the increment of the function when the argument moves from point P_0 to P_1 is accurately equal to the differential of the function corresponding to some intermediate point P_c on the straight line joining P_0 and P_1.*

This is *Lagrange's theorem (mean value theorem)* for a function of several independent variables. In particular, when P_0 and P_1 lie on Ox, we get Lagrange's theorem for a function of one independent variable (see Sec. 65).

II. The remainder term of Taylor's formula (*):

$$R = \frac{1}{(n+1)!}\left[\frac{\partial}{\partial x}(x - x_0) + \frac{\partial}{\partial y}(y - y_0)\right]^{n+1} f_{\substack{x=\xi \\ y=\eta}} = \frac{1}{(n+1)!} d^{n+1} f(P_c)$$

gives the error of the approximation

$$\Delta f(P_0) \approx df(P_0) + \frac{1}{2!} d^2 f(P_0) + \cdots + \frac{1}{n!} d^n f(P_0).$$

This error is an infinitesimal of order not lower than $(n + 1)$ with respect to $\varrho = P_0P_1 \to 0$. In particular, the approximation

$$\Delta f(P_0) \approx df(P_0)$$

is characterized by an error with an order of smallness of at least 2 with respect to ϱ.

The differential of a twice differentiable function differs from its increment by an infinitesimal of order at least 2.

If the remainder term R_n tends to zero at every point of some neighbourhood of the point P_0 as n increases indefinitely, $\lim_{n\to\infty} R_n = 0$, the function $f(P)$ can be written in this neighbourhood as an *in-*

finite power series, called the Taylor series for the function $f(P)$ (see Sec. 129):

$$f(P) = f(P_0) + df(P_0) + \frac{1}{2!} d^2 f(P_0) + \cdots$$

The Taylor series for the function $f(x, y)$ may be written in full as

$$f(x, y) = f(x_0, y_0) + [f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0)] + \\ + \frac{1}{2!}[f''_{x^2}(x_0, y_0)(x - x_0)^2 + 2f''_{xy}(x_0, y_0)(x - x_0)(y - y_0) + \\ + f''_{y^2}(x_0, y_0)(y - y_0)^2] + \cdots$$

Definition. A function which can be written as a convergent Taylor series in a certain domain is said to be analytic in this domain.

The conditions in which a function is analytic are: (1) it must be "infinitely differentiable", so that it has partial derivatives of any order; (2) the remainder term of Taylor's formula of the n-th order must tend to zero for any point of the domain when n increases indefinitely.

156. Extrema. Necessary Conditions.

Definition. **A point P_0 is called an *extremal* (*maximum* or *minimum*) *point* of a function $z = f(P)$ if the value taken by the function at the point is respectively greater or less than all the values taken in some neighbourhood of P_0.**

The value $f(P)_0$ is called the extremum (maximum or minimum) of the function. We can also say that $f(P)$ has an extremum (or attains an extremum), at the point $P = P_0$.

As in the case of a function of one variable (Sec. 63), the definitions of extremal points assume strict inequalities: *P_0 is a maximum (or minimum) point of the function $f(P)$ if a neighbourhood of P_0 exists such that we have $f(P) < f(P_0)$ (or $f(P) > f(P_0)$) for any point P of the neighbourhood.* We might provide for the possibility of there being points in any sufficiently small neighbourhood of P_0 at which the values of the function are equal to $f(P_0)$. In this case we should have to write $f(P) \leqslant f(P_0)$ (or $f(P) \geqslant f(P_0)$) instead of the strict inequalities.

We shall not change our definition, however, and the cases when $f(P) = f(P_0)$ will be set aside since they are rarely encountered outside the general theory. Each such case can easily be given special consideration.

It may be observed that, by virtue of our definition, an extremal point necessarily lies inside the domain of definition of the

function, so that the function is defined in some neighbourhood (even though small) of this point. The form of the surface representing the function in the neighbourhood of extrema is shown in Fig. 17.

We shall first of all establish the necessary conditions for a function $z = f(x, y)$ to attain an extremum at a point $P_0(x_0, y_0)$.

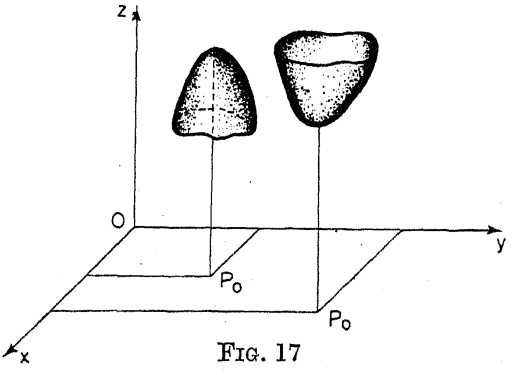

FIG. 17

We shall assume that we are dealing with functions of two independent variables that have continuous partial derivatives of the first order.

NECESSARY TEST FOR AN EXTREMUM. **If a function $z = f(x, y)$ be differentiable at a point $P_0(x_0, y_0)$ and attains an extremum at this point, its partial derivatives vanish at the point (its total differential vanishes):**

$$\left(\frac{\partial z}{\partial x}\right)_{P_0} = 0, \quad \left(\frac{\partial z}{\partial y}\right)_{P_0} = 0,$$

i. e.

$$(dz)_{P_0} = \left(\frac{\partial z}{\partial x}\right)_{P_0} dx + \left(\frac{\partial z}{\partial y}\right)_{P_0} dy = 0.$$

Proof. Let $z = f(x, y)$ have an extremum at $P_0(x_0, y_0)$.

By the definition, $z = f(x, y)$, regarded as a function of x only and with constant $y = y_0$, attains an extremum at $x = x_0$. We know that the necessary condition for this is that the derivative of $f(x, y)$ vanishes at $x = x_0$, i.e.

$$\frac{\partial f(x_0, y_0)}{\partial x} = 0, \quad \text{or} \quad \left(\frac{\partial z}{\partial x}\right)_{\substack{x = x_0 \\ y = y_0}} = 0.$$

Similarly, $z = f(x, y)$, regarded as a function of y only and with constant $x = x_0$, attains an extremum at $y = y_0$, i.e.

$$\frac{\partial f(x_0, y_0)}{\partial y} = 0, \quad \text{or} \quad \left(\frac{\partial z}{\partial y}\right)_{\substack{x = x_0 \\ y = y_0}} = 0.$$

This is what we had to prove.

A point $P_0(x_0, y_0)$ whose co-ordinates cause both partial derivatives of a function $z = f(x, y)$ to vanish is termed a stationary point of the function.

The above condition is not sufficient, however: examples may readily be found of functions having no extremum at a stationary point. Let us take the function $z = xy$. Its partial derivatives vanish at the origin, yet it has no extremum for $x = 0$, $y = 0$. In fact, whilst it vanishes at the origin, it has positive values (in the first and third quadrants) as well as negative values (in the second and fourth quadrants) in any neighbourhood of the origin, i.e. zero is neither the greatest nor the least value of $z = xy$ in any circle with centre at the point $P_0(0, 0)$.

The equation of the tangent plane (see Sec. 146) to the surface $z = f(x, y)$:

$$z - z_0 = \left(\frac{\partial z}{\partial x}\right)_{P_0}(x - x_0) + \left(\frac{\partial z}{\partial y}\right)_{P_0}(y - y_0)$$

becomes for a stationary point $P_0(x_0, y_0)$ of $z = f(x, y)$:

$$z = z_0.$$

Thus the *necessary condition for a differentiable function* $z = f(x, y)$ *to attain an extremum at a point* $P_0(x_0, y_0)$ *implies geometrically that the tangent plane to the surface, i.e. to the graph of the function at the corresponding point, is parallel to the plane of the independent variables.*

If P_0 is in fact an extremal point, the tangent plane *does not intersect the surface* in a neighbourhood of the point of contact, but lies either *above* it (in the case if a maximum) or *below* it (in the case of a minimum); whilst if P_0 is a stationary but not an extremal point, the tangent plane *cuts* the surface in the neighbourhood of the point of contact. For instance, the tangent plane to the hyperbolic paraboloid $z = xy$ at the origin coincides with the Oxy plane, whilst the surface in the neighbourhood of the point of contact $M_0(0, 0, 0)$ lies on both sides and not just on one side of the tangent plane. This plane both touches the surface at $M_0(0, 0, 0)$ and cuts it at this point.

A point of a surface $z = f(x, y)$ corresponding to a stationary but not extremal point $P(x, y)$ plays to some extent the same role as a point of inflexion of a plane curve.

If P_0 is a "non-strict" extremal point of a function $z = f(x, y)$, the plane $z = z_0$ tangential to the surface at the corresponding point M_0 touches the

surface (in the neighbourhood of M_0) at an infinity of points other than M_0. These points usually form a curve lying in the plane $z = z_0$.

It follows from the necessary condition for an extremum that a differentiable function $f(x, y)$ can only have an extremum at points of the Oxy plane whose co-ordinates satisfy the equations

$$f'_x(x, y) = 0, \quad f'_y(x, y) = 0. \qquad (*)$$

We thus arrive at the rule:

To find the values of the independent variables at which a differentiable function $z = f(x, y)$ can have an extremum we must equate the partial derivatives with respect to x and y to zero and find the real roots of system (*) of two equations with two unknowns.

We obtain pairs of values of x and y as co-ordinates of the stationary points i.e. of possible extremal points of the function. The sufficient conditions for an extremum, which will be dealt with in Sec. 158, enable us to indicate which of the stationary points are extremal points and which are not; and we can also see whether an extremum is a maximum or a minimum. It is sometimes possible to discover the nature of a stationary point without having recourse to the sufficient conditions, in which case the working may be simplified. For instance, if it follows directly from the conditions of the problem that the function in question has a maximum or minimum at some point and at the same time system (*) is only satisfied at one point (i.e. for one pair of values of x, y), it is clear that this point must be the required extremal point of the function. In addition, it is possible to make use of special peculiarities of the given function and to draw conclusions on the basis of these regarding the nature of the stationary point.

Example. Let us find the extrema of the function

$$z = 2 + 2x + 4y - x^2 - y^2.$$

We equate its partial derivatives to zero:

$$\frac{\partial z}{\partial x} = 2 - 2x = 0, \quad \frac{\partial z}{\partial y} = 4 - 2y = 0.$$

These equations give: $x = 1, \quad y = 2.$

Our function thus has only the one stationary point $P(1, 2)$. The corresponding value of the function is $z = 7$, which we shall show to be a maximum. We rewrite the function as

$$z = -(x - 1)^2 - (y - 2)^2 + 7,$$

whence

$$z - 7 = -[(x - 1)^2 + (y - 2)^2].$$

It is clear from this equation that $z - 7$ cannot be a positive number, i.e. z does not exceed 7:

$$z - 7 \leqslant 0, \quad z \leqslant 7.$$

We conclude from this that $z = 7$ is not only a maximum of the function but is its greatest value throughout the Oxy plane.

Thus $P(1, 2)$ is the maximum point. This is completely obvious geometrically, since the graph of $z = 2 + 2x + 4y - x^2 - y^2$ is the paraboloid of revolution with axis parallel to Oz and directed towards the negative side of Oz, the vertex being at $M(1, 2, 7)$.

We note finally that a continuous function of two variables can have extrema at points where the function is non-differentiable (corresponding to "spikes" of the surface representing the function). For instance, $z = \sqrt{x^2 + y^2}$ obviously has a minimum at the origin equal to zero, although it is not differentiable at this point. Consequently, if we are considering continuous functions in general and not just differentiable functions, we must say that:

The stationary points or the points at which the function is non-differentiable are possible extremal points.

Such points are sometimes described as "critical."

We can establish in precisely the same way necessary conditions for an extremum of a differentiable function of n independent variables

$$u = f(x, y, \ldots, t).$$

The extremal points of a differentiable function of n independent variables must in fact be stationary points of the function, i.e. points whose co-ordinates lead to the vanishing of all n partial derivatives of the function. The systems of values of $x, y, \ldots, t$ at which the function $u = f(x, y, \ldots, t)$ attains its extrema are found among the solutions of the system of n equations with n unknowns:

$$f'_x(x, y, \ldots, t) = 0, \quad f'_y(x, y, \ldots, t) = 0, \ldots, \quad f'_t(x, y, \ldots, t) = 0.$$

157. Problems on Absolutely Greatest and Least Values. Suppose we want to find the absolutely greatest (least) value of a function $z = f(x, y)$ in some *closed* domain. If this value is attained by the function inside the domain, it is evidently an extremum. But it may happen that the absolutely greatest (least) value is taken by the function at a point lying *on the boundary* of the domain. Even in the case when the function is defined in a neighbourhood of this boundary point, it is possible for the point not to be extremal. In fact, suppose we take $z = xy$ in the domain $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant$

$\leqslant 1$. It is positive everywhere for $0 < x \leqslant 1$, $0 < y \leqslant 1$ and vanishes for $x = 0$ and for $y = 0$. Consequently this function takes its least value on two sides of the boundary of the domain. But no point of the boundary is extremal, since $z = xy$ takes negative values in the second and fourth quadrants.

The above leads to the following rule.

RULE FOR FINDING THE ABSOLUTELY GREATEST OR LEAST VALUE. **To find the absolutely greatest or least taken by a function $z = f(x, y)$ in a closed domain we must find all the maxima or minima* of the function contained in the domain together with the greatest or least values on the boundary of the domain. The greatest (least) of all these numbers will in fact be the required absolutely greatest (least) value.**

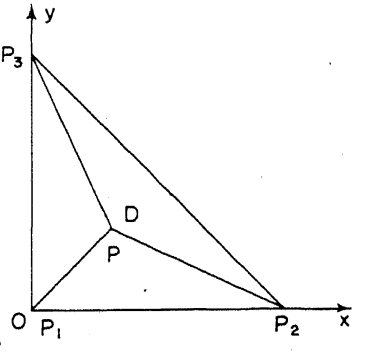

FIG. 18

Example. Let us find the point on the Oxy plane such that the sum of the squares of its distances from the three points $P_1(0, 0), P_2(1, 0), P_3(0, 1)$ has its least value, and the point in the triangle with vertices at P_1, P_2, P_3, such that the sum of the squares of its distances from the vertices has its greatest value (Fig. 18).

Let $P(x, y)$ be any point of the plane. The sum z of the squares of its distances from the given points is given by

$$z = x^2 + y^2 + (x - 1)^2 + y^2 + x^2 + (y - 1)^2,$$

or

$$z = 3x^2 + 3y^2 - 2x - 2y + 2.$$

The first part of the problem amounts to finding the least value of this function throughout the plane, and the second part to finding the greatest value of the function on condition that $P(x, y)$ belongs to the closed domain D bounded by triangle $P_1P_2P_3$.

We find the extrema of $z = 3x^2 + 3y^2 - 2x - 2y + 2$. The equations

$$\frac{\partial z}{\partial x} = 6x - 2 = 0 \quad \text{and} \quad \frac{\partial z}{\partial y} = 6y - 2 = 0$$

give us

$$x = \tfrac{1}{3}, \quad y = \tfrac{1}{3}.$$

* Instead of all the extrema of the function we can simply take the values at all the critical points inside the domain.

Thus there exists only one stationary point $P(\frac{1}{3}, \frac{1}{3})$. The function has no greatest value throughout the plane, since it is clear that points P exist for which the sum in question is greater than any previously assigned number. And since it is obvious on the other hand that this sum must have a least value, the stationary point $P(\frac{1}{3}, \frac{1}{3})$ must be the point at which the function attains its least value ($= 1\frac{1}{3}$). (Though it is easy to verify in the present case, just as in the example of Sec. 156, that $P(\frac{1}{3}, \frac{1}{3})$ is in fact a minimal point of the function throughout the plane, and is therefore the point required in the problem.) As a matter of fact, $P(\frac{1}{3}, \frac{1}{3})$ is the centre of gravity of the triangle with vertices P_1, P_2, P_3.

In view of the fact that our function has no maximum, its greatest value in domain D is the greatest of its values on the boundary, i.e. on the sides of the triangle.

We have $y = 0$ on side P_1P_2, i.e.

$$z = 3x^2 - 2x + 2;$$

this function has its greatest value ($= 3$) in the interval $[0, 1]$ at $x = 1$, i.e. at P_2.

We have $x = 0$ on side P_1P_3, i.e.

$$z = 3y^2 - 2y + 2;$$

the greatest value of this function in the interval $[0, 1]$ is also equal to 3 and occurs at $y = 1$, i.e. at P_3.

Finally, $x + y = 1$ on side P_2P_3, i.e.

$$z = 3x^2 + 3(1 - x)^2 - 2x - 2(1 - x) + 2 = 6x^2 - 6x + 3;$$

this function attains its greatest value ($= 3$) in the interval $[0, 1]$ at $x = 0$ and at $x = 1$, i.e. the greatest value on side P_2P_3 of function z is obtained at the same points P_2 and P_3. Thus there are two points P_2 and P_3 satisfying the requirements of the second part of the problem. These are the points such that the sum of the squares of their distances from the vertices of the triangle has its greatest value.

158. Sufficient Conditions for an Extremum. If function $f(x, y)$ has an extremum at the point $P_0(x_0, y_0)$, there exists a ϱ-neighbourhood of P_0 (i.e. a circle of radius ϱ with centre at P_0) such that we have for any point $P(x, y)$ of the neighbourhood:

$$f(P) < f(P_0) \quad \text{(in the case of a maximum)}$$

or

$$f(P) > f(P_0) \quad \text{(in the case of a minimum).}$$

On setting $x - x_0 = h$, $y - y_0 = k$, we can rewrite these inequalities as

$$f(x_0 + h, y_0 + k) < f(x_0, y_0) \quad \text{or} \quad f(x_0 + h, y_0 + k) > f(x_0, y_0)$$

for any h and k subject to the condition $0 < \sqrt{h^2 + k^2} < \varrho$.

Conversely, if there exists a $\varrho > 0$ such that, given any h and k satisfying the inequality $\sqrt{h^2 + k^2} < \varrho$, one of the above inequalities is satisfied, the function $f(x, y)$ has an extremum at $P_0(x_0, y_0)$ (a maximum with the first and a minimum with the second inequality).

On the other hand, if the inequality $f(x_0 + h, y_0 + k) < f(x_0, y_0)$ holds for certain h and k with *any* number ϱ (i.e. for certain points of the ϱ-neighbourhood of P_0), and $f(x_0 + h, y_0 + k) > f(x_0, y_0)$ for certain h and k, $f(x, y)$ has not got an extremum at $P_0(x_0, y_0)$ since it takes values both greater and less than $f(x_0, y_0)$ in any neighbourhood of P_0.

The question of the existence or absence of an extremum of $f(x, y)$ at a given point $P_0(x_0, y_0)$ is therefore answered by the sign of the difference

$$\Delta f = f(x_0 + h, y_0 + k) - f(x_0, y_0)$$

in a neighbourhood of P_0, i.e. for h and k subject to the condition $0 < \sqrt{h^2 + k^2} < \varrho$.

We now assume that $f(x, y)$ has continuous partial derivatives up to and including the second order at P_0. The difference Δf can now be transformed by Taylor's formula (Sec. 155):

$$\begin{aligned} f(x_0 + h, y_0 + k) - f(x_0, y_0) &= df(P_0) + \tfrac{1}{2} d^2 f(P_c) \\ &= \left(\frac{\partial f}{\partial x}\right)_{P_0} h + \left(\frac{\partial f}{\partial y}\right)_{P_0} k + \\ &+ \tfrac{1}{2}\left[\left(\frac{\partial^2 f}{\partial x^2}\right)_{P_c} h^2 + 2\left(\frac{\partial^2 f}{\partial x\, \partial y}\right)_{P_c} hk + \left(\frac{\partial^2 f}{\partial y^2}\right)_{P_c} k^2\right]; \end{aligned}$$

the point P_c has co-ordinates $x = x_0 + \theta h$, $y = y_0 + \theta k$, where $0 < \theta < 1$.

Since an extremum can only occur at a stationary point of the function, we can assume beforehand that

$$\left(\frac{\partial f}{\partial x}\right)_{P_0} = \left(\frac{\partial f}{\partial y}\right)_{P_0} = 0.$$

We shall suppose that not all the second order partial derivatives vanish at P_0. We put:

$$\left(\frac{\partial^2 f}{\partial x^2}\right)_{P_c} = \frac{\partial^2 f(x_0 + \theta h,\, y_0 + \theta k)}{\partial x^2} = \frac{\partial^2 f(x_0, y_0)}{\partial x^2} + \varepsilon_1,$$

$$\left(\frac{\partial^2 f}{\partial x\, \partial y}\right)_{P_c} = \frac{\partial^2 f(x_0 + \theta h,\, y_0 + \theta k)}{\partial x\, \partial y} = \frac{\partial^2 f(x_0, y_0)}{\partial x\, \partial y} + \varepsilon_2,$$

$$\left(\frac{\partial^2 f}{\partial y^2}\right)_{P_c} = \frac{\partial^2 f(x_0 + \theta h,\, y_0 + \theta k)}{\partial y^2} = \frac{\partial^2 f(x_0, y_0)}{\partial y^2} + \varepsilon_3.$$

In view of the assumed continuity of the second partial derivatives, ε_1, ε_2, and ε_3 tend to zero as $h \to 0$, $k \to 0$, i.e. as $\varrho \to 0$.

We simplify the writing by using the further notation:

$$\frac{\partial^2 f(x_0, y_0)}{\partial x^2} = A, \quad \frac{\partial^2 f(x_0, y_0)}{\partial x\, \partial y} = B, \quad \frac{\partial^2 f(x_0, y_0)}{\partial y^2} = C.$$

Now,

$$\begin{aligned} \Delta f &= f(x_0 + h, y_0 + k) - f(x_0, y_0) \\ &= \tfrac{1}{2}\,[(Ah^2 + 2Bhk + Ck^2) + \varepsilon], \end{aligned}$$

where

$$\varepsilon = \varepsilon_1 h^2 + 2\varepsilon_2 hk + \varepsilon_3 k^2.$$

As $\varrho \to 0$, ε is an infinitesimal of higher order than $Ah^2 + 2Bhk + Ck^2$, provided this last expression does not vanish for real h and k. The expression is in fact of the second order with respect to $\sqrt{h^2 + k^2}$, whilst ε is an infinitesimal of higher order than 2, since ε_1, ε_2, ε_3 are infinitesimals. But the sum of two infinitesimals of different orders has the same sign as the lower order infinitesimal for sufficiently small absolute values. Thus for sufficiently small ϱ, the sign of the difference Δf is the same as the sign of $Ah^2 + 2Bh + Ck^2$.

We observe that this last expression can be rewritten, on the assumption that A does not vanish, as

$$\begin{aligned} Ah^2 + 2Bhk + Ck^2 &= \frac{1}{A}(A^2h^2 + 2ABhk + B^2k^2 - B^2k^2 + ACk^2) \\ &= \frac{1}{A}[(Ah + Bk)^2 - k^2(B^2 - AC)]. \qquad (*) \end{aligned}$$

On the basis of this expression we can now state two sufficient conditions.

SUFFICIENT CONDITIONS FOR THE EXISTENCE OF AN EXTREMUM. **If P_0 is a stationary point of the function $f(x, y)$, and**

$$B^2 - AC < 0, \quad \text{i.e.} \quad \left(\frac{\partial^2 f}{\partial x\, \partial y}\right)^2_{P_0} - \left(\frac{\partial^2 f}{\partial x^2}\right)_{P_0}\left(\frac{\partial^2 f}{\partial y^2}\right)_{P_0} < 0 \qquad (**)$$

$f(x, y)$ has an extremum at P_0: a maximum if $A < 0$ $(C < 0)$, a minimum if $A > 0$ $(C > 0)$.

Proof. If condition (**) is satisfied, A and C differ from zero and have the same sign, since otherwise the expression $B^2 - AC$ would be positive; as is clear from formula (*), the expression $Ah^2 + 2Bhk + Ck^2$ has a constant sign the same as that of A for any h and k. Hence, if $A < 0$, $\Delta < 0$ for any sufficiently small h and k, and we have a maximum at the point P_0; whilst if $A > 0$, $\Delta > 0$ for any sufficiently small h and k, and we have a minimum at P_0.

SUFFICIENT CONDITIONS FOR THE ABSENCE OF AN EXTREMUM. **If P_0 is a stationary point of the function $f(x, y)$ and**

$$B^2 - AC > 0, \quad \text{i.e.} \quad \left(\frac{\partial^2 f}{\partial x\, \partial y}\right)^2_{P_0} - \left(\frac{\partial^2 f}{\partial x^2}\right)_{P_0}\left(\frac{\partial^2 f}{\partial y^2}\right)_{P_0} > 0, \quad (***)$$

$f(x, y)$ has no extremum at P_0.

Proof. If condition (***) is satisfied and $A \neq 0$, it is clear from formula (*) that, given sufficiently small h and k, $Ah^2 + 2Bhk + Ck^2$ can take negative as well as positive values*. Consequently there exist points P in any neighbourhood of P_0 at which the difference $\Delta f > 0$, i.e. at which the function is greater than at P_0 itself, and points P at which $\Delta f < 0$, i.e. at which the function is less than its value at P_0. This means that there is no extremum at P_0. Whereas if $A = 0$, then $B \neq 0$ (since otherwise we should have $B^2 - AC = 0$), and it is easily seen that the expression

$$Ah^2 + 2Bhk + Ck^2 = 2Bk^2\left(\frac{h}{k} + \frac{C}{2B}\right)$$

can take both positive and negative values for sufficiently small h and k, which implies, as above, that P_0 is not an extremal point.

If** $$B^2 - AC = 0,$$

* For instance, if $A > 0$, this expression has negative values for all h and k satisfying the condition $Ah + Bk = 0$, i.e. along the whole of the straight line $A(x - x_0) + B(y - y_0) = 0$, whilst with $k = 0$ and all h, i.e. along the whole of the straight line $y = y_0$, it has positive values.

** This equation still holds in the case excluded above, when all the partial derivatives of the second order vanish: $A = B = C = 0$.

nothing definite can be said about the nature of the stationary point without further investigation—which must in fact be omitted in the present course. *There may or may not be an extremal point in this case.*

The above can be summarized in the following rule.

RULE FOR FINDING EXTREMA. **To find the extremal points and values of a twice differentiable function $z = f(x, y)$ in a given domain, we must**

(1) Equate the partial derivatives to zero:

$$\frac{\partial z}{\partial x} = 0, \quad \frac{\partial z}{\partial y} = 0$$

and find the real roots of this system of two equations with two unknowns. Each pair of roots defines a stationary point of the function. We have to take those stationary points that lie in the given domain.

(2) Evaluate the expression

$$B^2 - AC,$$

where $A = \partial^2 z/\partial x^2$, $B = \partial^2 z/\partial x\, \partial y$, $C = \partial^2 z/\partial y^2$, at each stationary point.

Here:

(a) if $B^2 - AC < 0$, we have an extremum: a maximum with $A < 0$ (and $C < 0$), a minimum with $A > 0$ (and $C > 0$);

(b) if $B^2 - AC > 0$, there is no extremum;

(c) if $B^2 - AC = 0$, we have the indeterminate case, requiring special investigation;

(3) Work out the extremal values of the function; this is done by substituting the co-ordinates of the extremal points in the expression for the function.

Example 1. We have for the function $z = 2 + 2x + 4y - x^2 - y^2$ (see the example of Sec. 156):

$$A = \frac{\partial^2 z}{\partial x^2} = -2, \quad B = \frac{\partial^2 z}{\partial x\, \partial y} = 0, \quad C = \frac{\partial^2 z}{\partial y^2} = -2.$$

Thus

$$B^2 - AC = -4 < 0,$$

and the stationary point (1, 2) is a maximum, since

$$A = C = -2 < 0.$$

Example 2. We have for $z = xy$:

$$A = 0, \quad B = 1, \quad C = 0$$

and

$$B^2 - AC = +1 > 0.$$

The stationary point (0, 0) does not give an extremum.

Example 3. We have at the stationary point (0, 0) of the function $z = x^3y^3$: $B^2 - AC = 0$. Similarly, at the stationary point (0, 0) of $z = x^4y^4$, we again have $B^2 - AC = 0$. It may easily be verified by considering the function directly that $z = x^3y^3$ has no extremum, whilst $z = x^4y^4$ has a minimum at the origin.

159. Conditional Extrema. Up to now we have posed the problem of the extrema of a function of two independent variables as follows: given the function, it is required to find its extremal points belonging either to the whole or part of its domain of definition. When the argument of the function, i.e. the point in the plane of the independent variables, varies, no restrictions are imposed on the variation by this statement of the problem, apart from the fact that the point must stay in the domain concerned. Such extrema are described as unconditional or free.

Problems are often encountered, however, in which the so-called conditional extrema are required.

Let the function $z = f(x, y)$ be given together with its domain of definition D. Let L be a given curve in domain D and let the extrema of $f(x, y)$ be required corresponding to points of L only. These are in fact the conditional extrema of $z = f(x, y)$ on the curve L.

Definition. **The value of a function $f(x, y)$ at some point $P_0(x_0, y_0)$ of a curve L is said to be a *conditional extremum* of the function if it is greater or less than all the other values of $f(x, y)$ at points of L belonging to some neighbourhood of the point $P_0(x_0, y_0)$.**

The free movement of the argument of the function—the point $P(x, y)$—is here limited to the curve L, and the extremum is distinguished by comparing the given value of the function only with the adjacent values corresponding to points of L.

If the equation of L is

$$\varphi(x, y) = 0,$$

the problem of finding the conditional extrema of $z = f(x, y)$ on L can be stated analytically as follows:

To find the extrema of the function $z = f(x, y)$ in the domain D with the condition that $\varphi(x, y) = 0$.

Thus it is no longer possible to regard x and y as *independent* variables when finding the conditional extrema of $f(x, y)$; they are connected by the relationship $\varphi(x, y) = 0$. The left-hand side of this relationship is called the connecting function between variables x and y, whilst $\varphi(x, y) = 0$ is termed the connecting equation.

We take a simple example to illustrate the difference between the unconditional and conditional extrema of a function of two variables. The unconditional maximum of the function $z = \sqrt{1 - x^2 - y^2}$

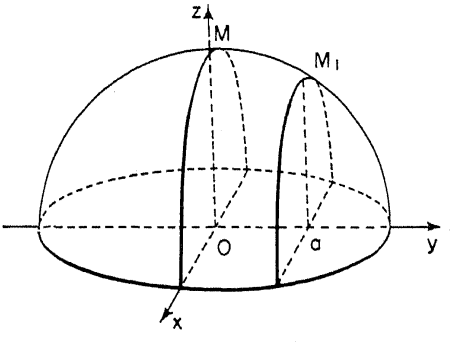

FIG. 19

(the graph of which is the upper hemisphere, Fig. 19) is equal to unity; it is attained at the point (0, 0) and corresponds to the vertex M (north pole) of the sphere. Yet the conditional maximum of

$$z = \sqrt{1 - x^2 - y^2},$$

say on the straight line $y - a = 0$, $0 < a < 1$, is obviously equal to $\sqrt{1 - a^2}$; it is attained at the point $(0, a)$ and corresponds to the vertex M_1 of the semi-circle in which the hemisphere and the plane $y = a$ intersect.

It may be shown that the search for a *conditional extremum of a function* can be reduced to a search for an *unconditional extremum of some other function.*

Suppose that, for the values of x and y in question, the equation $\varphi(x, y) = 0$ defines y as a single-valued differentiable function: $y = \psi(x)$. On substituting $\psi(x)$ for y in the expression for the function $z = f(x, y)$, we obtain a function of the single variable x:

$$z = f[x, \psi(x)] = \Phi(x).$$

The extrema of the function $\Phi(x)$ are in fact the required conditional extrema of $z = f(x, y)$ with connecting equation

$\varphi(x, y) = 0$. The search for the extrema of $z = \Phi(x)$ is carried out in accordance with the familiar rules (see Sec. 67 and 70).

This method of reducing a conditional to a free extremal problem is not always convenient in practice, since it requires the actual solution of the equation $\varphi(x, y) = 0$ with respect to one of the variables.

Another method may be mentioned which does not require an implicit function to be reduced to an explicit, and which is readily extended to the case of a function of any number of independent variables.

Let $P_0(x_0, y_0)$ be a point at which $z = f(x, y)$ has a conditional extremum. On taking as y the function $\psi(x)$ defined by the connecting equation $\varphi(x, y) = 0$, the derivative with respect to x of $f(x, y)$ at the point P_0 may be seen to be necessarily zero, or what amounts to the same thing, the differential of $f(x, y)$ at P_0 must be equal to zero. Since the expression for the differential of a function does not depend on the nature of its arguments, we can write

$$(df)_{P_0} = (f'_x dx + f'_y dy)_{P_0} = 0.$$

At the same time, we have from the connecting equation:

$$(d\varphi)_{P_0} = (\varphi'_x dx + \varphi'_y dy)_{P_0} = 0.$$

On multiplying the latter equation by an as yet undetermined factor λ (independent of x and y) and adding term by term to the first equation, we get

$$(f'_x + \lambda\varphi'_x)_{P_0} dx + (f'_y + \lambda\varphi'_y)_{P_0} dy = 0.$$

Let λ be chosen to that

$$(f'_y + \lambda\varphi'_y)_{P_0} = 0;$$

then we also have

$$(f'_x + \lambda\varphi'_x)_{P_0} = 0.$$

These two equations express the familiar necessary conditions for a (free) extremum at the point $P_0(x_0, y_0)$ of the function

$$f(x, y) + \lambda\varphi(x, y).$$

It follows from this that a point at which $f(x, y)$ has a conditional extremum with $\varphi(x, y) = 0$ must be a stationary point of the function

$$F(x, y) = f(x, y) + \lambda\varphi(x, y),$$

where λ is a coefficient.

RULE FOR FINDING CONDITIONAL EXTREMA. **To find the points at which a function $z = f(x, y)$ can have conditional extrema with the connecting equation $\varphi(x, y) = 0$, we must form the auxiliary function**

$$F(x, y) = f(x, y) + \lambda\varphi(x, y)$$

and find its stationary points. The equations

$$F'_x(x, y) = f'_x(x, y) + \lambda\varphi'_x(x, y) = 0,$$

$$F'_y(x, y) = f'_y(x, y) + \lambda\varphi'_y(x, y) = 0,$$

$$\varphi(x, y) = 0$$

form a system of three equations giving the value of λ and the coordinates x, y of possible extremal points.

This last method of reducing a conditional to an unconditional extremal problem is known as the method of **Lagrange multipliers.**

We shall not concern ourselves with the means for completing the solution of the problem, i.e. with discovering whether or not the points found by Lagrange's method are in fact points at which conditional extrema occur (and if so, in what sense). This question can usually be answered by considerations of a subsidiary kind, as indicated in Sec. 156.

Lagrange's method can be extended to functions of any number of independent variables.

Suppose we want to find the (conditional) extrema of a function of n variables

$$u = f(x, y, z, \ldots, t),$$

when variables $x, y, z, \ldots, t$ are not independent but are subject to m ($m < n$) *connecting equations:*

$$\varphi_1(x, y, z, \ldots, t) = 0,$$

$$\varphi_2(x, y, z, \ldots, t) = 0,$$

$$\ldots\ldots\ldots\ldots\ldots$$

$$\varphi_m(x, y, z, \ldots, t) = 0.$$

By using arguments precisely similar to the above for the case of functions of two variables, we arrive at the rule:

To find the points which may be the required conditional extremal points, we have to form the auxiliary function

$$F(x, y, z, \ldots, t) = f(x, y, z, \ldots, t) + \lambda_1 \varphi_1(x, y, z, \ldots, t) + \cdots \\ \cdots + \lambda_m \varphi_m(x, y, z, \ldots, t)$$

and find its stationary points.

The equations

$$F'_x = f'_x + \lambda_1 \varphi'_{1x} + \cdots + \lambda_m \varphi'_{mx} = 0,$$
$$F'_y = f'_y + \lambda_1 \varphi'_{1y} + \cdots + \lambda_m \varphi'_{my} = 0,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$F'_t = f'_t + \lambda_1 \varphi'_{1t} + \cdots + \lambda_m \varphi'_{mt} = 0$$

together with the m connecting equations $\varphi_1 = 0, \varphi_2 = 0, \ldots, \varphi_m = 0$, *form a system of* $n + m$ *equations from which we can find the values of the m parameters* $\lambda_1, \lambda_2, \ldots, \lambda_m$, *and the n co-ordinates* $x, y, z, \ldots, t$ *of possible extremal points.*

Example 1. Let us find the distance from a given point to a given straight line.

We shall assume for simplicity that the given point lies at the origin (this can always be arranged for by parallel displacement of the axes) and that the given straight line is in the Oxy plane. Let the equation of the straight line be $Ax + By + C = 0$.

The distance z of any point $P(x, y)$ of the Oxy plane from the point $P_0(0, 0)$ is given by $z = \sqrt{x^2 + y^2}$. We have to find the least value of this function on condition that $P(x, y)$ lies on the straight line $Ax + By + C = 0$; this least value is clearly a minimum of the function. We have therefore arrived at the problem of the conditional minimum of $z = \sqrt{x^2 + y^2}$ with connecting equation $Ax + By + C = 0$.

We form the auxiliary function

$$F(x, y) = \sqrt{x^2 + y^2} + \lambda(Ax + By + C)$$

and seek its stationary points. We have:

$$F'_x = \frac{x}{\sqrt{x^2 + y^2}} + \lambda A = 0, \quad F'_y = \frac{y}{\sqrt{x^2 + y^2}} + \lambda B = 0.$$

We add to these equations the connecting equation

$$Ax + By + C = 0$$

and obtain a system of three equations for finding the three quantities λ, x, y.

On multiplying the first equation by B, the second by A and subtracting one from the other, we get

$$Ay - Bx = 0.$$

We notice that this is the equation of a straight line passing through the origin and perpendicular to the given line.

We find from the equation obtained and the connecting equation a unique system of values for x and y. On substituting these in the expression for the function, we obtain the familiar formula of analytic geometry:

$$z = \frac{|C|}{\sqrt{A^2 + B^2}}.$$

The fact that this value of z is in fact the required conditional minimum follows from the obvious absence of other extrema of the function and the uniqueness of the values of x and y satisfying the equations

$$F'_x = 0 \quad \text{and} \quad F'_y = 0.$$

Example 2. Let us find the greatest value of the n-th root of the product of n positive numbers $x, y, z, \ldots, t$, on condition that their sum is equal to a given number A.

The problem amounts to finding the conditional maximum of the function

$$u = \sqrt[n]{xy \ldots t}$$

in the domain $x > 0$, $y > 0$, $\ldots$, $t > 0$, with the connecting equation

$$x + y + \cdots + t - A = 0.$$

We take the auxiliary function

$$F(x, y, \ldots, t) = \sqrt[n]{xy \ldots t} + \lambda(x + y + \cdots + t - A).$$

The necessary conditions for an extremum give

$$F'_x = \frac{1}{n} \frac{\sqrt[n]{xy \ldots t}\, y \ldots t}{xy \ldots t} + \lambda = \frac{1}{n} \frac{u}{x} + \lambda = 0 \quad \text{or} \quad u = -n\lambda x,$$

$$F'_y = \frac{1}{n} \frac{u}{y} + \lambda = 0 \quad \text{or} \quad u = -n\lambda y,$$

. . . . . .

$$F'_t = \frac{1}{n} \frac{u}{t} + \lambda = 0 \quad \text{or} \quad u = -n\lambda t.$$

On comparing all these equations, we find that

$$x = y = \cdots = t,$$

the common value of $x, y, \ldots, t$ being A/n by virtue of the connecting equation.

As above, we conclude from the uniqueness of the stationary point of function F and the actual nature of the problem that the numbers

$$x = \frac{A}{n}, \quad y = \frac{A}{n}, \quad \ldots, \quad t = \frac{A}{n}$$

give the greatest value of $u = (xy \ldots t)^{1/n}$. This value is equal to A/n, i.e.

$$\sqrt[n]{xy \ldots t} \leqslant \frac{x + y + \cdots + t}{n}.$$

The quantity on the left-hand side is termed the *geometric mean of numbers* $x, y, \ldots, t$, whilst that on the right is the *arithmetic mean of the numbers.*

Thus *the geometric mean of positive numbers is always less than their arithmetic mean, or is equal to the arithmetic mean if all the numbers are equal.*

An elementary proof of the particular case $n = 2$ can easily be given either analytically or geometrically.

2. Elements of Vector Analysis

160. Vector Function of a Scalar Argument. *Differentiation.* We assume that the reader is familiar with vector algebra (at least from a course of analytic geometry) so that we can pass over this and turn directly to the elementary problems of vector analysis.

Only free vectors are considered here, i.e. vectors which can be reckoned equal if they have equal moduli and the same direction (free vectors are the same if they differ only in their initial points).

I. VECTOR FUNCTION. HODOGRAPH. A vector, like a scalar, can be either *constant* or *variable.* A vector is variable if it takes different vector "values"; it is constant if it keeps the same vector "value." A variable vector changes its position in space whereas a constant vector does not. Velocity is an obvious example of a variable vector, provided the motion is not uniform along a straight line.

Definition. **A vector A is said to be a (single-valued) *function of a point* P in some domain D of space (or of the plane), if for every position of the point in D there is a corresponding definite "value" of vector A, i.e. definite values of its absolute value (modulus) and direction (or of all its projections).**

We indicate the fact that vector $\mathbf{A}$ is a function of the point P by using the notation $\mathbf{A} = \mathbf{A}(P)$, and describe $\mathbf{A}$ as a vector function of point P.

We shall confine ourselves here to the case when the point P varies along an axis; in this case the position of P is given by a

single scalar argument, which is often time (denoted by t). Thus, for every value of t in some interval there is a corresponding definite vector $\mathbf{A}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$, the projections x, y, z of which on the co-ordinate axes are functions of t.

We shall picture vector $\mathbf{A}(t)$ as always starting from the origin; as t varies, the end of the vector $\mathbf{A}(t)$, with co-ordinates $x(t)$, $y(t)$, $z(t)$ will now describe some curve L, the equation of which is given by the following parametric equations*:

$$x = x(t), \quad y = y(t), \quad z = z(t).$$

Since vector $\mathbf{A}(t)$ is precisely the same thing as the radius vector $\mathbf{r}$ of the point M on L, the curve can be specified by the vector equation

$$\mathbf{r} = x(t)\,\mathrm{i} + y(t)\,\mathrm{j} + z(t)\,\mathrm{k}$$

(or $\mathbf{r} = x(t)\mathbf{i} + y(t)\mathbf{j}$ when L is a plane curve), where $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ denote as usual the *base vectors*.

Definition. **The curve L described by the end of vector $\mathbf{A}$ is called the *hodograph* of the vector function $\mathbf{r} = \mathbf{A}(t)$, whilst the origin is the *pole of the hodograph*.**

The trajectory of a moving point is the hodograph of the radius vector of the point as a function of time. In general, every curve is the hodograph of the radius vector of a point of it as a function of the corresponding parameter. If only the absolute value of $\mathbf{A}(t)$ changes, the hodograph will be a straight line starting from the pole; whereas if only the direction changes, the hodograph is a curve lying on the sphere with centre at the pole and radius equal to the constant absolute value of $\mathbf{A}(t)$.

Analytic concepts such as limit, continuity etc. can be brought in for vector functions $\mathbf{A}(t)$ of a scalar argument t just as in the case of scalar functions.

Definition. **Vector $\mathbf{A}$ is said to be the *limit of vector function* $\mathbf{A}(t)$ *as* $t \to t_0$:**

$$\lim_{t \to t_0} \mathbf{A}(t) = \mathbf{A},$$

if, given any positive ε, a corresponding positive δ can be found such that, for all t satisfying the condition $0 < |t - t_0| < \delta$ we have

$$|\mathbf{A} - \mathbf{A}(t)| < \varepsilon.$$

* The reader should not be disturbed by the fact that the same notation is used for the function and for its dependence on the argument (or arguments). This situation will often be encountered in future.

Definition. **A function $\mathbf{A}(t)$ is said to be *continuous* at $t = t_0$ if it is defined in the neighbourhood of t_0 and**

$$\lim_{t \to t_0} \mathbf{A}(t) = \mathbf{A}(t_0).$$

As in the case of scalar functions, a different form of definition can be given, based on the concept of increment. We give the argument t an increment Δt and find the corresponding increment $\Delta \mathbf{A}(t)$:

$$\Delta \mathbf{A}(t) = \mathbf{A}(t + \Delta t) - \mathbf{A}(t).$$

The geometric meaning of this increment is as follows. Suppose a point M of the hodograph L of function $\mathbf{A}(t)$ corresponds to the

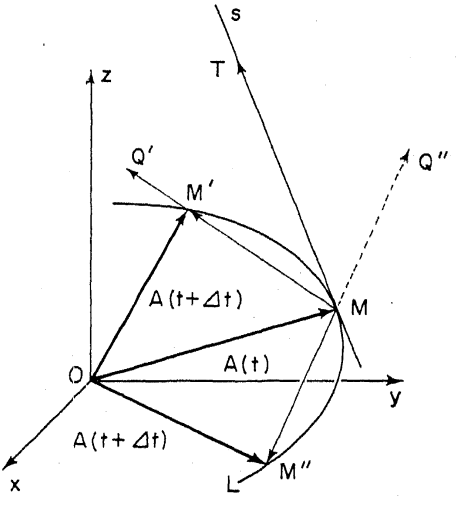

FIG. 20

value t of the argument, whilst M' is the point corresponding to $t + \Delta t$ if $\Delta t > 0$, or M'' if $\Delta t < 0$ (Fig. 20). Then $\Delta \mathbf{A}(t)$ is either the vector $\mathbf{MM'}$ or $\mathbf{MM''}$.

If $\Delta \mathbf{A}(t) \to 0$ (or what amounts to the same thing, $|\Delta \mathbf{A}(t)| \to 0$) *as* $\Delta t \to 0$, *the vector function* $\mathbf{A}(t)$ *is said to be continuous at the point.*

Continuity of the vector implies continuity of the hodograph, and conversely, the vector is continuous if its hodograph is a continuous curve.

II. VECTOR DERIVATIVE. We now establish the concept of the rate of change of a vector function $\mathbf{A}(t)$ at a given value t.

We shall assume, not only that the hodograph is a continuous curve, but also that it has a tangent at every point.

We take the ratio $\Delta \mathbf{A}(t)/\Delta t$, which is termed the average rate of change of vector $\mathbf{A}(t)$ in the interval $(t, t + \Delta t)$. This ratio is

represented geometrically by the vector $\mathbf{MQ}'$ (or $\mathbf{MQ}''$) directed towards the side of increasing t. For, if $\Delta t < 0$, the vector $\Delta \mathbf{A}(t) = \mathbf{MM}''$ is in the opposite direction, but on division by the negative number Δt we get vector $\mathbf{MQ}''$, which is directed, like vector $\mathbf{MQ} = \Delta \mathbf{A}(t)/\Delta t$, $\Delta t > 0$, to the part of the hodograph corresponding to increase of parameter t. As $\Delta t \to 0$, let the chord MM' (or MM'') of the curve L tend to the position of the straight line Ms, which is said to be tangential to L at M.

Definition. **The limit**

$$\lim_{\Delta t \to 0} \frac{\Delta \mathbf{A}(t)}{\Delta t}$$

is called the *derivative of vector function* $\mathbf{A}(t)$ *with respect to the scalar argument* t.

The derivative is seen to be the *vector* $\mathbf{MT}$, *tangential to the hodograph of vector* $\mathbf{A}(t)$ and in a direction corresponding to increase of t. The vector derivative is written as $\mathbf{A}'(t)$ or $d\mathbf{A}(t)/dt$:

$$\mathbf{A}'(t) = \mathbf{MT};$$

this vector *expresses the rate of change of vector function* $\mathbf{A}(t)$ *at the point* t.

Let us consider in particular the case when L is the trajectory of the motion of a point M and t is time; $\mathbf{r} = \mathbf{A}(t)$ is now called the *equation of motion*, and L the *hodograph of the motion*, whilst the vector derivative $\mathbf{v} = d\mathbf{r}/dt$ is the *velocity of the motion*. The velocity is therefore the vector defined above, tangential to the trajectory at the corresponding point.

If the function $\mathbf{A}(t)$ has constant absolute value, say equal to unity: $|\mathbf{A}(t)| = 1$, its derivative $\mathbf{A}'(t)$ is a vector *perpendicular to vector* $\mathbf{A}(t)$. For, the hodograph lies on a sphere, so that $\mathbf{A}'(t)$, being a vector tangential to the hodograph, is perpendicular to the radius vector of the corresponding point of the sphere, which is in fact the given vector $\mathbf{A}(t)$. This proposition can be briefly stated as:

The derivative of a unit vector is perpendicular to it.

III. Differentiation. Expansion of the velocity. By starting from the rules for the arithmetic operations on vectors we can develop, as in the case of scalar functions, a differential calculus for vector functions of a scalar argument. In view of the fact that the rules of arithmetic hold for vectors, the ordinary rules of differentiation remain in force for vector functions.

For instance, we can show that

$$\frac{d}{dt}(\mathbf{A}_1(t) \times \mathbf{A}_2(t)) = \mathbf{A}_1(t) \times \frac{d\mathbf{A}_2(t)}{dt} + \frac{d\mathbf{A}_1(t)}{dt} \times \mathbf{A}_2(t)$$

(though it is not possible to permute the factors in this expression), or that

$$\frac{d\mathbf{A}}{ds} = \frac{d\mathbf{A}}{ds'}\frac{ds'}{ds}$$

(the rule for differentiation of a function of a function) etc.

Similarly, by expanding a vector in the base vectors, the operation of differentiation on vector functions can be reduced to the corresponding operation on scalar functions. Let

$$\mathbf{A}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + r(z)\,\mathbf{k}.$$

We have

$$\Delta\mathbf{A}(t) = \Delta x(t)\,\mathbf{i} + \Delta y(t)\,\mathbf{j} + \Delta z(t)\,\mathbf{k},$$

$$\frac{\Delta\mathbf{A}(t)}{\Delta t} = \frac{\Delta x(t)}{\Delta t}\mathbf{i} + \frac{\Delta y(t)}{\Delta t}\mathbf{j} + \frac{\Delta z(t)}{\Delta t}\mathbf{k};$$

on passing to the limit as $\Delta t \to 0$ in accordance with the usual rule, we get

$$\mathbf{A}'(t) = x'(t)\,\mathbf{i} + y'(t)\,\mathbf{j} + z'(t)\,\mathbf{k}, \qquad (*)$$

i.e. the *base vector expansion of the vector derivative of* $\mathbf{A}(t)$. It will be seen that this expansion has the same form as if we differentiated $\mathbf{A}(t)$ simply as the sum $x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$ with constant $\mathbf{i}, \mathbf{j}, \mathbf{k}$. This leads to the rule:

To differentiate a vector, we simply differentiate its projections on to the co-ordinate axes.

As regards the geometrical meaning of expansion (*), we note first of all that the definitions of *tangent* (which we have used above) and *arc length* can be carried over without change (see Sec. 164) to the case of spatial curves. If we write s for the arc length of L, given by the equation $\mathbf{r} = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$, where x, y, z are the current co-ordinates on L, we have

$$ds = \sqrt{dx^2 + dy^2 + dz^2} = \sqrt{x'^2(t) + y'^2(t) + z'^2(t)}\,dt.$$

Here ds is the length of the corresponding segment of the tangent to the curve. We find for the direction cosines $\cos\alpha$, $\cos\beta$, $\cos\gamma$ of the tangent:

$$\cos\alpha = \frac{dx}{ds}, \quad \cos\beta = \frac{dy}{ds}, \quad \cos\gamma = \frac{dz}{ds}$$

(see also Sec. 164). It follows from these expressions that vector (*) has the same direction cosines as the tangent, i.e. it is in the direction of the tangent; we discovered this above by a rather different method.

Since, by expansion (*),

$$|\mathbf{A}'(t)| = \sqrt{x'^2(t) + y'^2(t) + z'^2(t)},$$

we have

$$|\mathbf{A}'(t)| = \frac{ds}{dt}, \qquad (**)$$

and we can now completely characterize the derivative:

***The derivative of a vector function of a scalar argument* is a vector tangential to the hodograph of the given vector and equal in absolute value to the derivative of the arc length of the hodograph with respect to the argument.**

In particular, when $\mathbf{r} = \mathbf{A}(t)$ is an equation of motion, we find that the *absolute value of the velocity is equal to the derivative of the path with respect to time*: $|\mathbf{v}| = ds/dt$; in the case of rectilinear motion, the *velocity vector degenerates to a scalar*, which we called in Sec. 12 the speed of the motion and defined as the derivative ds/dt.

If we take the arc length s of the hodograph as the argument t, the absolute value of the vector derivative is always equal to unity, as is clear from formula (**). The derivative of the radius vector of a point of an arc with respect to the arc length is therefore a vector tangential to the arc and of length unity. In this case the parametric equations of the curve:

$$x = x(s), \quad y = y(s), \quad z = z(s) \quad (\text{or} \quad \mathbf{r} = x(s)\,\mathbf{i} + y(s)\,\mathbf{j} + z(s)\,\mathbf{k})$$

are termed the *natural equations*.

Like any vector, $\mathbf{r} = \mathbf{A}(t)$ can be written as the product of its absolute value and a unit vector in the same direction:

$$\mathbf{A}(t) = |\mathbf{A}(t)|\,\mathbf{r}_1.$$

Hence,

$$\frac{d\mathbf{A}(t)}{dt} = \frac{d|\mathbf{A}(t)|}{dt}\,\mathbf{r}_1 + |\mathbf{A}(t)|\,\frac{d\mathbf{r}_1}{dt}.$$

The first term on the right-hand side is a vector with the same direction as $\mathbf{r}_1$, i.e. the direction of the given vector $\mathbf{A}(t)$; the absolute value of the first term is equal to the derivative of the absolute value of the given vector. The second term on the right is a vector having the direction of vector $d\mathbf{r}_1/dt$, i.e. a direction perpendicular to the given vector $\mathbf{A}(t)$.

The formula thus gives the *resolution of the vector derivative along the direction of the radius vector of the hodograph and along the perpendicular direction.*

IV. SECOND DERIVATIVE. RESOLUTION OF THE ACCELERATION. We arrive by successive differentiation at higher order derivatives of a vector function with respect to a scalar argument. We shall dwell on the second derivative.

The second derivative $\mathbf{A}''(t)$ of a vector function $\mathbf{A}(t)$ can be obtained by twice differentiating the base vector expansion of $\mathbf{A}(t)$:

$$\mathbf{A}''(t) = x''(t)\,\mathbf{i} + y''(t)\,\mathbf{j} + z''(t)\,\mathbf{k}.$$

The second derivative can be expressed differently, however, if we start from the form

$$\mathbf{A}'(t) = |\mathbf{A}'(t)|\,\boldsymbol{\tau}_1,$$

where $\boldsymbol{\tau}_1$ is the unit vector tangential to the hodograph $\mathbf{r} = \mathbf{A}(t)$.

We find by differentiation:

$$\mathbf{A}''(t) = \frac{d\mathbf{A}'(t)}{dt} = \frac{d\,|\mathbf{A}'(t)|}{dt}\,\boldsymbol{\tau}_1 + |\mathbf{A}'(t)|\,\frac{d\boldsymbol{\tau}_1}{dt},$$

which gives the resolution of the second derivative along directions tangential to the hodograph and normal to it.

In the case of motion given by $\mathbf{r} = \mathbf{A}(t)$, the vector $\mathbf{w} = \mathbf{A}''(t) = \mathbf{v}'$ is called the *acceleration*, the first component $\mathbf{w}_t = [d\mathbf{A}'(t)/dt]\,\boldsymbol{\tau}_1$ being the tangential acceleration and the second component $\mathbf{w}_n = |\mathbf{A}'(t)|\,d\boldsymbol{\tau}_1/dt$ being the normal acceleration.

The coefficient of the unit tangential vector in the expression for the tangential acceleration is d^2s/dt^2, i.e. *the second derivative of the length of the trajectory with respect to time.*

As regards the normal acceleration $\mathbf{w}_n$, its expression can be written as

$$\mathbf{w}_n = |\mathbf{A}'(t)|\,\frac{d\boldsymbol{\tau}_1}{dt} = \frac{ds}{dt}\,\frac{d\boldsymbol{\tau}_1}{ds}\,\frac{ds}{dt} = \left(\frac{ds}{dt}\right)^2 \left|\frac{d\boldsymbol{\tau}_1}{ds}\right| \boldsymbol{\nu}_1,$$

where $\boldsymbol{\nu}_1$ is a unit vector having the same direction as $d\boldsymbol{\tau}_1/ds$, normal to the trajectory $\mathbf{r} = \mathbf{A}(t)$. We have further:

$$\left|\frac{d\boldsymbol{\tau}_1}{ds}\right| = \left|\frac{d\boldsymbol{\tau}_1}{ds'}\right| \left|\frac{ds'}{ds}\right|,$$

where s' is the arc length of the hodograph of vector $\boldsymbol{\tau}_1$, in view of which $|d\boldsymbol{\tau}_1/ds| = 1$, and, since $\boldsymbol{\tau}_1$ is a unit vector, ds' expresses

the infinitesimal angle of displacement $d\varphi$ (see Sec. 80) of the trajectory $\mathbf{r} = \mathbf{A}(t)$. For a spatial curve, as for a plane curve, the quantity $|d\varphi/ds|$ is taken as the special characteristic called *curvature:* $|d\varphi/ds| = 1/R$, where R is the *radius of curvature.* Thus

$$\mathbf{w}_n = \frac{|\mathbf{v}|^2}{R}\,\boldsymbol{\nu}_1.$$

It is clear from this that the *absolute value of the normal acceleration is equal to* $|\mathbf{v}|^2/R$, where $|\mathbf{v}|^2 = (ds/dt)^2$ is the square of the velocity, whilst R is the radius of curvature of the trajectory.

The acceleration vector $\mathbf{w}$ can be written as follows:

$$\mathbf{w} = \mathbf{w}_t + \mathbf{w}_n = \frac{d|\mathbf{v}|}{dt}\,\boldsymbol{\tau}_1 + \frac{|\mathbf{v}|^2}{R}\,\boldsymbol{\nu}_1 = \frac{d^2s}{dt^2}\,\boldsymbol{\tau}_1 + \frac{\left(\frac{ds}{dt}\right)}{R}\,\boldsymbol{\nu}_1.$$

In the case of rectilinear motion the acceleration vector has no normal component: *it degenerates to a scalar, equal to* d^2s/dt^2, which we referred to in Sec. 59 as the *acceleration of rectilinear motion.*

161. Gradient. Let a scalar function $z = f(P) = f(x, y)$ be defined in a given domain D of the plane referred to Cartesian co-ordinates Oxy and let it be *differentiable* (in other words, a *scalar field* is given in the domain D).

If the point P is displaced in the direction α characterized by the unit vector $\mathbf{e}_\alpha = \cos\alpha\,\mathbf{i} + \cos\beta\,\mathbf{j}$ ($\cos\beta = \sin\alpha$) (α is the angle between the direction and Ox), the rate of change of the function at P is given by the directional derivative

$$\frac{\partial z}{\partial \alpha} = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\cos\beta.$$

We take the vector with initial point at P whose projections on Ox and Oy are $\partial z/\partial x$ and $\partial z/\partial y$ respectively.

Since the projections of vector $\mathbf{e}_\alpha$ on Ox and Oy are equal to $\cos\alpha$ and $\sin\alpha$ respectively, the derivative $\partial z/\partial\alpha$ is equal to the scalar product of the vector just introduced and the unit vector $\mathbf{e}_\alpha$.

Definition. **The vector $(\partial z/\partial x)\mathbf{i} + (\partial z/\partial y)\mathbf{j}$, defined for the point $P(x, y)$, is termed the *gradient of the function* $z = f(P)$ (or *of the scalar field*) at the point P and is written as grad z (grad $f(P)$):**

$$\operatorname{grad} z = \frac{\partial z}{\partial x}\,\mathbf{i} + \frac{\partial z}{\partial y}\,\mathbf{j}.$$

Thus every point of a scalar field has *associated with it* a definite *vector*, in other words, a *scalar function* $f(P)$ in a domain D generates a *vector function* grad $f(P)$ in this domain.

We now apply the gradient to the local investigation of a function. We have

$$\frac{\partial z}{\partial \alpha} = \operatorname{grad} z \cdot \mathbf{e}_\alpha , \qquad (*)$$

i.e.

The derivative of a function with respect to a given direction is equal to the scalar product of the gradient of the function and the unit vector in this direction.

Since the scalar product is equal to the absolute value of one vector multiplied by the projection of the other on the direction of the first, we can also say that

The derivative of a function with respect to a given direction is equal to the projection of the gradient of the function at the given point on the direction of differentiation.

The expression (*) for the directional derivative is convenient in that it enables us to see the effect of the direction of differentiation on the derivative.

It is interesting in this connection to discover the actual direction of the gradient of a function at a given point.

THEOREM. **The gradient of a function is directed normally to the level curve of the function passing through the given point.**

Proof. We draw the level line passing through a given point P_0; the value of the function remains constant along it: $f(P) = \text{const}$. We take the derivative $\partial z/\partial s$ at the point P with respect to the direction of this curve; we know (Sec. 148) that it is equal to zero: $\partial z/\partial s = \operatorname{grad} z \cdot \mathbf{e}_s = 0$, where $\mathbf{e}_s$ is the unit vector tangential to the level line. But the scalar product is equal to zero (if grad $z \neq 0$) only when the vector factors are perpendicular to each other; thus grad z is perpendicular to $\mathbf{e}_s$, i.e. *is normal to the level line*. This is what we wanted to prove.

The absolute value of the gradient is equal to $\sqrt{z_x'^2 + z_y'^2}$, whilst its direction cosines are

$$\cos\alpha = \frac{z_x'}{\sqrt{z_x'^2 + z_y'^2}}, \quad \cos\beta = \sin\alpha = \frac{z_y'}{\sqrt{z_x'^2 + z_y'^2}}.$$

If we find the direction of the gradient from these expressions, we can also easily find the direction of the level curve of the function passing through the given point.

It follows at once from formula (*) that the greatest (absolute) value of the directional derivative is got when the direction of vector $\mathbf{e}_\alpha$ (i.e. the direction of differentiation) is the same as or opposite to that of grad z.

Hence *the direction of the gradient is the direction with respect to which the derivative* $\partial z/\partial \alpha$ *has its greatest* (absolute) *value, equal to* $\sqrt{z_x'^2 + z_y'^2}$. It follows from this that

The directional derivative of a function $z = f(x, y)$ at a point $P(x,y)$ takes its greatest (absolute) value, equal to $\sqrt{z_x'^2 + z_y'^2}$, when it is taken with respect to the normal to the level curve of $f(x, y)$ passing through the point $P(x, y)$.

(We observe that, if the function is increasing at the rate $\sqrt{z_x'^2 + z_y'^2}$ on one side of the normal, it is decreasing at the same rate on the other side.) The direction of the gradient is the direction of steepest rise of the surface $z = f(x, y)$ at the point in question.

The concept of gradient can be extended to functions of any number of independent variables; we confine ourselves here to three variables.

Definition. **The *gradient* grad $f(P)$ of the function $u = f(P) = f(x, y, z)$ is the vector whose projections on the axes Ox, Oy, Oz are $\partial u/\partial x$, $\partial u/\partial y$, $\partial u/\partial z$ respectively.**

As above, we have for the derivative with respect to a direction PN:

$$f'_{PN}(P) = \text{grad } f(P) \cdot \mathbf{e}_{PN},$$

where $\mathbf{e}_{PN}$ is the unit vector in the direction of PN. (Its projections on the axes are the direction cosines $\cos\alpha$, $\cos\beta$, $\cos\gamma$ of PN.)

We conclude from this that the directional derivative has its greatest value when it is taken with respect to the direction of the gradient. Since the derivative of $f(x, y, z)$ with respect to any direction tangential to a level surface is zero (Sec. 148), the gradient is directed perpendicularly to any straight line tangential to the level surface, i.e. to the tangent plane (see Sec. 166).

Such a direction is called a *normal to the surface.*

Thus *the gradient of a function of three independent variables at any point is normal to the level surface passing through this point,* i.e. *the direction of the normal to a level surface is the direction of greatest change of the function.*

The gradient of a function $u = f(P)$ of several independent variables is to some extent equivalent to the derivative of a function of one independent variable.

We take at a point $P(x, y, z)$ the vector whose projections on Ox, Oy, Oz are dx, dy, dz respectively (the displacement vector of point P). We write it symbolically as $\overrightarrow{dP}$. This notation is entirely natural: the vector $\overrightarrow{dP}$, like the differential of the independent variable, completely characterizes the displacement of P: its magnitude, equal to $\sqrt{dx^2 + dy^2 + dz^2}$, measures the distance by which P is displaced, whilst its direction gives the direction of displacement. The differential of $u = f(P) = f(x, y, z)$ can obviously be written as the scalar product

$$du = df(P) = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy + \frac{\partial u}{\partial z}\,dz = \operatorname{grad} f(P) \cdot \overrightarrow{dP}.$$

On writing symbolically $\overrightarrow{f'(P)}$ for $\operatorname{grad} f(P)$, we get a single term for the differential of $u = f(P)$:

$$du = \overrightarrow{f'(P)} \cdot \overrightarrow{dP}.$$

If P varies only along an axis, say Ox, the vector $\overrightarrow{dP}$ degenerates to a scalar dx, and vector $\overrightarrow{f'(P)}$ to the scalar $f'(x)$, the scalar product to the ordinary product, and the formula as a whole to the familiar expression for the differential of the function $f(x)$.

3. Curves. Surfaces

162. Plane Curves. We take a curve L in a plane given by the general equation

$$F(x, y) = 0,$$

which is not soluble for either co-ordinate. We can now apply the differential calculus for functions of two independent variables to investigating plane curves in the general case, and thus supplement what was said in Chapters III and IV regarding curves given by equations of the form

$$y = f(x).$$

Definition. If, in a neighbourhood of a point of a curve, one co-ordinate of the point can be expressed as a single-valued continuously differentiable function of the other co-ordinate, the point is said to be regular.

Let $M_0(x_0, y_0)$ be a regular point of a curve L. Then in the neighbourhood of M_0 the function $F(x, y)$ is such that the equation $F(x, y) = 0$ defines* $y = f(x)$ as a function of x which is single-valued and differentiable in the neighbourhood of the point $x = x_0$ and for which

$$f'(x_0) = -\frac{F'_x(x_0, y_0)}{F'_y(x_0, y_0)},$$

where $F'_y(x_0, y_0) \neq 0$.

* See the existence theorem for implicit functions in Sec. 151.

The equation of the tangent at $M_0(x_0, y_0)$ *to the curve* L *is*

$$y - y_0 = -\frac{F'_x(x_0, y_0)}{F'_y(x_0, y_0)}(x - x_0),$$

or

$$F'_x(x_0, y_0)(x - x_0) + F'_y(x_0, y_0)(y - y_0) = 0,$$

which may be written more briefly as

$$F'_x \cdot (x - x_0) + F'_y \cdot (y - y_0) = 0,$$

or

$$\frac{\partial F}{\partial x} \cdot (x - x_0) + \frac{\partial F}{\partial y} \cdot (y - y_0) = 0.$$

When writing the expression in this form it should be remembered that the partial derivatives are taken at the point of contact $M_0(x_0, y_0)$, i.e. with $x = x_0$, $y = y_0$.

The equation of the normal to curve L *at the same point is*

$$F'_y \cdot (x - x_0) - F'_x \cdot (y - y_0) = 0.$$

We have for the differential of the length of arc ds:

$$ds = \sqrt{1 + \frac{F'^2_x}{F'^2_y}}\, dx = \frac{\sqrt{F'^2_x + F'^2_y}}{F'_y}\, dx$$

and

$$ds = \frac{\sqrt{F'^2_x + F'^2_y}}{F'_x}\, dy.$$

The direction cosines $\cos\alpha$ and $\cos\beta\,(= \sin\alpha)$ of the tangent are thus given by

$$\cos\alpha = \frac{dx}{ds} = \frac{F'_y}{\sqrt{F'^2_x + F'^2_y}}, \quad \cos\beta = \frac{dy}{ds} = \frac{F'_x}{\sqrt{F'^2_x + F'^2_y}}.$$

Similarly, expressions can be found for the curvature and the angle between two intersecting curves with equations that cannot be solved with respect to either co-ordinate.

Let both partial derivatives vanish at the point $M_0(x_0, y_0)$ on curve L:

$$F'_x(x_0, y_0) = 0, \quad F'_y(x_0, y_0) = 0;$$

in this case the direction of the curve (i.e. of its tangent) cannot be found at M_0 by the above formula, which becomes meaningless.

Definition. If neither co-ordinate of a point of a curve can be expressed in the neighbourhood of some given point as a single-valued continuously differentiable function of the other co-ordinate, the given point is described as singular.

By the definition, the partial derivatives $\partial F/\partial x$ and $\partial F/\partial y$ must vanish at a singular point $M_0(x_0, y_0)$ of a curve $F(x, y) = 0$. This means that the co-ordinates x_0, y_0 of singular points of the curve L are to be sought among the real roots of the simultaneous equations

$$F(x, y) = 0, \quad F'_x(x, y) = 0, \quad F'_y(x, y) = 0.$$

Roots of this system of equations are not necessarily singular points, however. For instance, the point (0, 0) lies on the curve $y^3 - x^5 = 0$ and both partial derivatives vanish here:

$$\frac{\partial F}{\partial x} = -5x^4, \quad \frac{\partial F}{\partial y} = 3y^2.$$

At the same time, this is an ordinary point of the curve; the equation can in fact be written as $y = x^{5/3}$, (0, 0) being a point of inflexion with the tangent coinciding with Ox.

We shall not dwell on the *sufficient tests* for singular points and their classification. We shall merely mention that singular points include cusps (e.g. (0, 0) for the semicubical parabola $y^2 - x^3 = 0$ (see Sec. 21)) and nodes (e.g. (0, 0) for the folium of Descartes $x^3 + y^3 - 3axy = 0$ (see Sec. 75)).

163. The Envelope of a Family of Plane Curves. Let us consider a special geometrical problem which we have already encountered in separate examples and which has a significance for the theory of differential equations (see Chapter XIV). This problem is concerned with families (i.e. systems) of curves.

Definition. **The equation of a *family* of plane curves is an equation between the two current co-ordinates which depends on a number of arbitrary parameters and expresses a curve of the family for definite numerical values of the parameters.**

The family of curves is described as *one-*, *two-*, *three-parameter*, etc. according to the number of parameters appearing in the equation of the family.

The equation of a one-parameter family has the form

$$f(x, y, C) = 0,$$

where C is an arbitrary parameter; the equation of a two-parameter family is

$$f(x, y, C, D) = 0,$$

where C, D are arbitrary and independent parameters, etc.

Choosing values for the parameters implies distinguishing one particular curve of the family; on varying these parameters we pass in general to another curve of the family.

Thus the equation

$$(x - C)^2 + y^2 = 1$$

is the equation of a one-parameter family of circles of radius 1 with centre on Ox; the equation

$$(x - C)^2 + (y - D)^2 = 1$$

gives the two-parameter family of all circles of radius 1; whilst the equation with three parameters,

$$(x - C)^2 + (y - D)^2 = E^2$$

gives the three-parameter family of circles with any centre and radius, i.e. the family of all circles on the plane.

We shall only deal with *one-parameter* families of plane curves.

Definition. **If all the curves of a family are touched by the same curve and the latter is touched at every point by a curve of the family*, it is called the *envelope* of the family.**

In the cases usually encountered the envelope so to speak envelopes all the curves of the family whilst the latter in aggregate outline the envelope.

Example 1. The family

$$(x - C)^2 + y^2 = 1$$

of all circles of constant radius 1 with centres on Ox has an envelope made up of the two parallel straight lines $y = \pm a$ (Fig. 21) (see Sec. 45).

Example 2. The evolute of a curve is the envelope of the family of normals to the curve (see Sec. 82).

Example 3. In general, *any curve is the envelope of the family of all its tangents* (see Sec. 45).

* This proviso is made because it can happen say that all the curves of the family are touched by some curve only at individual points. Thus all the circles $x^2 + (y - C)^2 = C^2$ are touched by Ox only at the origin. We do not call Ox the envelope.

It is easy to point to a one-parameter family of curves which has no envelope. For instance, the family of parallel straight lines $y = x + C$ and the family of integral curves $y = \int f(x)\,dx$ have no envelopes.

The following theorem gives the method of finding the envelope, given the equation of a one-parameter family of curves.

THEOREM. **If curves of the family**

$$f(x, y, C) = 0 \qquad (*)$$

have no singular points, the curve defined by the system of two equations

$$f(x, y, C) = 0, \quad f'_C(x, y, C) = 0,$$

is the *envelope* of family (*) provided it also has no singular points.

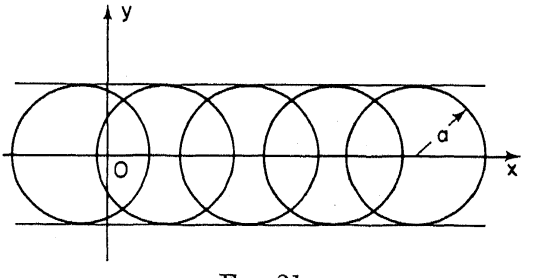

FIG. 21

Proof. Let us first suppose that family (*) has an envelope L. We shall express L by means of the parametric equations

$$x = \varphi(t), \quad y = \psi(t).$$

We choose these equations so as to satisfy the following condition: the value of parameter t at which a given point $M(x, y)$ of L is obtained coincides with the value of parameter C in equation (*) for which the curve of the family is obtained which touches envelope L at the point M. (Such a choice of parametric equations is always possible†). We can therefore put $t = C$ in the equation of L:

$$x = \varphi(C), \quad y = \psi(C). \qquad (**)$$

It is assumed by hypothesis that functions f, φ, ψ are differentiable and that $f'^2_x + f'^2_y \neq 0$, $\varphi'^2 + \psi'^2 \neq 0$. Since the envelope and a curve of the family have a common tangent at their common

† We shall not dwell on the proof of this.

point the slopes obtained for the tangent from equations (*) and (**) must be equal. We have from (*):

$$y' = -\frac{f'_x(x, y, C)}{f'_y(x, y, C)},$$

and from (**):

$$y' = \frac{\psi'(C)}{\varphi'(C)}.$$

Thus

$$f'_x(x, y, C)\,\varphi'(C) + f'_y(x, y, C)\,\psi'(C) = 0. \qquad (1)$$

We take the total derivative of function f with respect to C on the assumption that x and y are the co-ordinates of the point at which the curve and envelope touch, i.e. that $x = \varphi(C)$, $y = \psi(C)$.

We get

$$f'_x(x, y, C)\,\varphi'(C) + f'_y(x, y, C)\,\psi'(C) + f'_C(x, y, C) = 0. \qquad (2)$$

Comparison of (2) and (1) gives us

$$f'_C(x, y, C) = 0,$$

which must be satisfied by the co-ordinates of the point of contact, in other words, by the co-ordinates of a point of envelope L. The equations

$$f(x, y, C) = 0, \quad f'_C(x, y, C) = 0 \qquad (***)$$

define the current co-ordinates of the envelope as a function of parameter C, i.e. equations (**) must follow from them.

We now suppose that, conversely, x and y can be chosen from equations (***) as functions (**) of parameter C (if it is impossible to do this†, family (*) must have no envelope). It will be shown that curve (**) is the envelope of family (*).

In fact, on differentiating the identity

$$f[\varphi(C), \psi(C), C] = 0,$$

with respect to C, we get equation (2); since $f'_C(x, y, C) = 0$, we arrive at equation (1), from which it follows that curve (**) and

† For instance, for the family of straight lines $(C - 1)x - C^2y = 0$ we have $x - 2Cy = 0$; it is impossible to express x and y from these two equations as functions of parameter C. The family has in fact no envelope. Simple elimination of C from the two equations gives $x(\frac{1}{2}x - 2y) = 0$, i.e. the two straight lines $x = 0$ and $y = x/4$, belonging to the family (with $C = 0$ and $C = 2$).

the corresponding curve of family (*) have a common tangent, i.e. that curve (**) envelopes family (*). The theorem is proved.

If all the conditions of the theorem are not fulfilled, e.g. curves of family (*) have singular points, it is possible that the curve given by equations (**) obtained from system (***) is the locus of singular points of the family instead of being its envelope. This may be seen as follows: if curve (**) is the locus of singular points of curves (*), functions $\varphi(C)$ and $\psi(C)$ must satisfy system (***). For we have at singular points:

$$f'_x(x, y, C) = 0, \quad f'_y(x, y, C) = 0,$$

and these equations together with (2) yield $f'_C(x, y, C) = 0$.

Thus system (***) can define both the envelope of family (*) and the locus of singular points of curves of the family. In addition, as the example in the footnote on p. 102 has shown, the system can simply define individual curves of the family. The equations of all these curves are got from system (***) by eliminating parameter C.

Definition. **The curve given by the equation $\varphi(x, y) = 0$, obtained by eliminating parameter C from the equation of the family $f(x, y, C) = 0$ and the equation $f'_C(x, y, C) = 0$ is termed the *discriminant curve* of the family.**

Thus *the discriminant curve can include the envelope, the locus of singular points and separate curves of the family.* Having found the discriminant curve, we have to investigate directly the type of curve which it represents.

Example 1. Given the family of circles

$$(x - C)^2 + y^2 = 1$$

(Fig. 21), we find on differentiating with respect to C:

$$-2(x - C) = 0.$$

On combining this with the original equation we get the equation of the envelope:

$$y = \pm 1.$$

Example 2. We take the following family of straight lines (see Sec. 45, Fig. 56, Part I). We draw any straight line from the point $F(\frac{1}{2}p, 0)$ and erect the perpendicular to it at its intersection with Oy. The system of all such perpendiculars is the family of straight lines in question. Its equation will be

$$y = -\frac{1}{C}x - \frac{1}{2}Cp,$$

where C is the slope of the straight line through F. We get by differentiating with respect to C:

$$\frac{1}{C^2}x - \frac{1}{2}p = 0 \quad \text{or} \quad 2x = C^2 p.$$

Substitution in the equation of the family gives

$$y = \frac{-4x}{2\sqrt{\frac{2x}{p}}}, \quad \text{whence} \quad y^2 = 2px;$$

we have obtained the equation of a parabola. We have thus proved the assertion on which we based one of the methods of Sec. 45 for tracing parabolas.

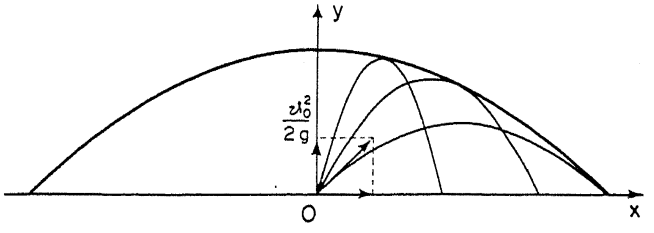

FIG. 22

Example 3. Let us consider the family of trajectories of a projectile fired with initial velocity v_0 at different angles α to the horizontal. We shall first of all find the equation of the family.

We direct the co-ordinate axes along the horizontal and vertical and locate the origin at the firing point (Fig. 22). We shall regard the projectile as a particle and neglect air resistance.

Let v_{0x} and v_{0y} be the projections of the initial velocity v_0 on the axes:

$$v_{0x} = v_0 \cos\alpha, \quad v_{0y} = v_0 \sin\alpha.$$

The projections of the velocity at the instant t can now be written as

$$v_x = v_0 \cos\alpha, \qquad v_y = v_0 \sin\alpha - gt.$$

Hence $dx = v_0 \cos\alpha \, dt$ and $dy = (v_0 \sin\alpha - gt)\, dt$

i.e. $x = v_0 \cos\alpha \cdot t$ and $y = v_0 \sin\alpha \cdot t - \frac{gt^2}{2}$.

This is the parametric equation of the trajectory. Elimination of t gives us the equation of a parabola:

$$y = \tan\alpha \cdot x - \frac{g}{2v_0^2 \cos^2\alpha} x^2.$$

The trajectory is therefore a parabola. The equation of the family can be written as

$$y = Cx - \frac{g}{2v_0^2}(1 + C^2)x^2,$$

where the parameter $C = \tan\alpha$ is the slope of the trajectory at the origin O.

Differentiation with respect to C gives

$$x - \frac{g}{v_0^2} Cx^2 = 0, \quad \text{whence} \quad C = \frac{v_0^2}{gx}.$$

The equation of the envelope will be

$$y = \frac{v_0^2}{g} - \frac{g}{2v_0^2} x^2 - \frac{v_0^2}{2g} = \frac{v_0^2}{2g} - \frac{g}{2v_0^2} x^2.$$

The envelope of the trajectories is therefore also a parabola (Fig. 22). It is called the parabola of safety. The paraboloid of revolution formed by rotating the parabola of safety about its axis divides space into two parts: points of the part lying between the paraboloid and the earth can be hit with the given initial velocity v_0, points of the part lying outside the paraboloid cannot be hit at the initial velocity v_0 no matter what the angle α of inclination of the gun to the horizontal.

Example 4. The discriminant curve has merely been the envelope in the above examples. An example will now be mentioned when the discriminant curve includes both the envelope and the locus of singular points of curves of the family. We take the family of semicubical parabolas

$$y^2 = (x - C)^3.$$

Differentiation with respect to C gives

$$-3(x - C)^2 = 0;$$

on eliminating the parameter we get the equation of the discriminant curve:

$$y = 0.$$

The discriminant curve—Ox (Fig. 23)—is here the locus of the cusps of the semicubical parabolas and at the same time is the envelope of the family, although it differs from the cases usually encountered in that it in no sense "envelopes" the system.

The discriminant curve of the family of semicubical parabolas (Fig. 24)

$$(y - C)^2 = x^3$$

is Oy; it is only the locus of the cusps, and not the envelope. The present family has no envelope.

164. Spatial Curves. The Helix.

I. TANGENT LINES AND NORMAL PLANES. A curve in space can be defined by three parametric equations for the co-ordinates:

$$x = x(t), \quad y = y(t), \quad z = z(t)$$

or by one vector equation:

$$\mathbf{r} = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}$$

(we shall assume that functions $x(t)$, $y(t)$, $z(t)$ are differentiable).

When the parameter varies in an interval, the end of the radius vector with co-ordinates (projections) x, y, z describes a given curve.

Definition. The **tangent line** *to a curve at the point $M_0(x_0, y_0, z_0)$ is the line passing through M_0 and occupying the "limiting position"**

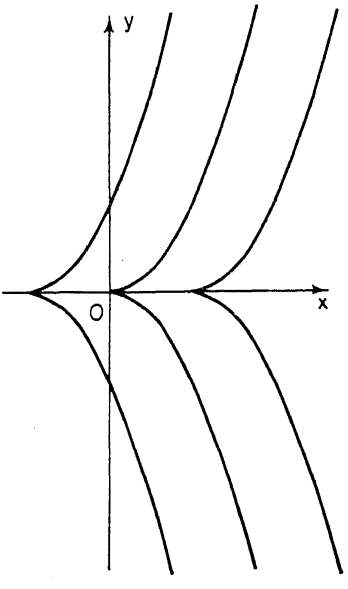

FIG. 23

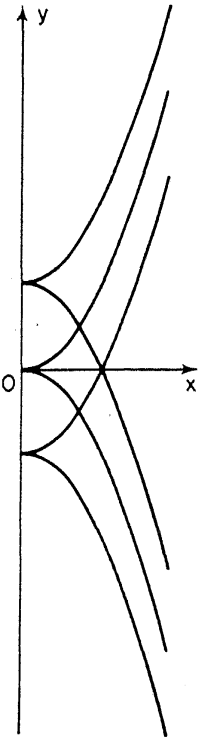

FIG. 24

M_0T of the secant through M_0 and another point M' of the curve when M' tends to M_0.

We actually made use of this definition in Sec. 160; it was found there that the vector derivative

$$\frac{d\mathbf{r}}{dt} = x'(t)\,\mathbf{i} + y'(t)\,\mathbf{j} + z'(t)\,\mathbf{k}$$

lies on the tangent. It follows from this that the slopes of the tan-

* This means that angle TM_0M' tends to zero as point M' tends to M_0.

gent must be equal to $x'(t)$, $y'(t)$, $z'(t)$ respectively, its equation being expressible as

$$\frac{x - x_0}{x'(t_0)} = \frac{y - y_0}{y'(t_0)} = \frac{z - z_0}{z'(t_0)},$$

where $x_0 = x(t_0)$, $y_0 = y(t_0)$, $z_0 = z(t_0)$.

These equations may easily be deduced directly. The equations of the secant through $M_0(x_0, y_0, z_0)$ and $M'(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z)$ will be: $(x - x_0)/\Delta x = (y - y_0)/\Delta y = (z - z_0)/\Delta z$, whence, on dividing all the denominators by the increment Δt of the parameter and passing to the limit as $\Delta t \to 0$, we obtain the above equations.

The direction cosines of the tangent at $M_0(x_0, y_0, z_0)$ are given by (see Sec. 160):

$$\cos\alpha = \frac{x'(t_0)}{\sqrt{x'^2(t_0) + y'^2(t_0) + z'^2(t_0)}},$$

$$\cos\beta = \frac{y'(t_0)}{\sqrt{x'^2(t_0) + y'^2(t_0) + z'^2(t_0)}},$$

$$\cos\gamma = \frac{z'(t_0)}{\sqrt{x'^2(t_0) + y'^2(t_0) + z'^2(t_0)}}$$

or

$$\cos\alpha = \frac{dx}{\sqrt{dx^2 + dy^2 + dz^2}}, \quad \cos\beta = \frac{dy}{\sqrt{dx^2 + dy^2 + dz^2}},$$

$$\cos\gamma = \frac{dz}{\sqrt{dx^2 + dy^2 + dz^2}}.$$

Definition. A straight line perpendicular to the tangent and passing through the point of contact is called a normal *to the curve at the point* $M_0(x_0, y_0, z_0)$.

The curve obviously has an infinite number of normals at each point: all these lie in a plane, perpendicular to the tangent and passing through the point of contact.

Definition. The plane perpendicular to the tangent line at its point of contact with the curve is called the normal plane *at that point.*

Since the normal plane is a plane perpendicular to the straight line $(x - x_0)/x'(t_0) = (y - y_0)/y'(t_0) = (z - z_0)/z'(t_0)$ and passing through the point $M_0(x_0, y_0, z_0)$, we know from analytic geometry that its equation must be

$$x'(t_0)(x - x_0) + y'(t_0)(y - y_0) + z'(t_0)(z - z_0) = 0.$$

II. LENGTH OF ARC. Let $M(x, y, z)$ and $M'(x + \Delta x, y + \Delta y, z + \Delta z)$ be points on the curve. We take as the length of arc Δs between these two points the quantity equivalent (as $\Delta x \to 0$, $\Delta y \to 0$, $\Delta z \to 0$) to the length of chord between M and M'. Since the length of the chord is equal to $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$, we have

$$\frac{\Delta s}{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}} \to 1.$$

The expression for the differential ds of the length of arc follows from this. We have from the last relationship:

$$\frac{\frac{\Delta s}{\Delta t}}{\sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2 + \left(\frac{\Delta z}{\Delta t}\right)^2}} \to 1,$$

where Δt is the increment of the parameter corresponding to the co-ordinate increments Δx, Δy, Δz. On passing to the limit as $\Delta t \to 0$, we get

$$\frac{\frac{ds}{dt}}{\sqrt{x'^2 + y'^2 + z'^2}} = 1,$$

i.e. $ds = \sqrt{x'^2 + y'^2 + z'^2}\, dt$ or $ds = \sqrt{dx^2 + dy^2 + dz^2}$.

But this shows that the arc differential at point $M(x, y, z)$ consists of an infinitesimal segment of the tangent at this point. We therefore have the propositon, as in the plane case:

An infinitesimal relative error is obtained (in measuring the arc length) *if we replace the infinitesimal arc of a spatial curve by the corresponding segment of its tangent.*

We again arrive, with the aid of the differential of the arc length, at the formulae given in Sec. 160 for the direction cosines of the tangent:

$$\cos\alpha = \frac{dx}{ds}, \quad \cos\beta = \frac{dy}{ds}, \quad \cos\gamma = \frac{dz}{ds}.$$

These expressions show the obvious fact that the projections of the segment ds of tangent on the axes are the corresponding co-ordinate increments dx, dy, dz.

Having found the expression for the differential of the arc length, the total length can be found by integration*:

$$L = \int_{t_1}^{t_2} \sqrt{x'^2 + y'^2 + z'^2}\, dt,$$

where t_1 and t_2 are the values of parameter t corresponding to the initial and final points of the arc.

The formula for the arc length can be written symbolically as

$$L = \int_{(A)}^{(B)} ds = \int_{(A)}^{(B)} \sqrt{dx^2 + dy^2 + dz^2},$$

where (A) and (B) denote the beginning and end of the arc.

III. The helix. A commonly encountered example of a spatial curve is the cylindrical helix. It can be defined by the following kinematic conditions: a point M moves with constant linear velocity v_1 over a circle representing the section normal to the axis of a right circular cylinder of radius a, whilst the circle itself gradually moves along the cylinder at the same time with constant velocity v_2. The point M now describes a helix (Fig. 25). In the case when M moves along the circle anticlockwise looking in the direction of the axial motion, the helix is said to be right-handed (as in Fig. 25). In the contrary case the helix is left-handed.

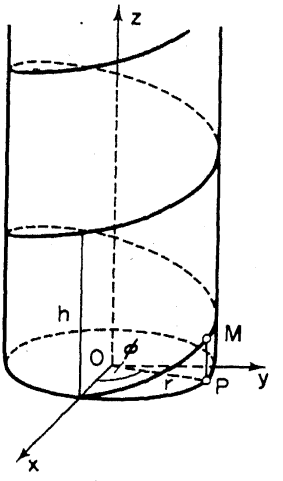

Fig. 25

We shall deduce the equation of the helix on the supposition that the axis of the cylinder is Oz and that the axial motion is in the direction of positive Oz, the point M being at (1, 0, 0) (Fig. 25) at the instant $t = 0$.

The angular velocity of the rotation of M is equal to v_1/a, so that its abscissa and ordinate at the instant t are evidently $x = a \cos(v_1/a)\, t$, $y = a \sin(v_1/a)\, t$. The z co-ordinate is equal to the height to which the point has been raised at instant t, i.e. $z = v_2 t$.

* This definition of the length of arc is equivalent to taking this length as the limit of the perimeter of the inscribed polygon (i.e. the polygon consisting of chords) when the number of sides increases indefinitely and the greatest side tends to zero.

If, instead of taking t as parameter we take the polar angle φ of the projection P of point M on the Oxy plane, the equation of the helix becomes

$$x = a\cos\varphi, \quad y = a\sin\varphi, \quad z = c\varphi, \quad \text{where} \quad c = \frac{v_2}{v_1}a.$$

This is the right-handed helix; the only difference as regards the left-handed helix is in the sign in front of coefficient c. This fact is bound up with our use of a *co-ordinate system* which is itself right-handed.

On eliminating parameter φ, we can write the helix as

$$x = a\cos\frac{z}{c}, \quad y = a\sin\frac{z}{c}.$$

The first equation is the equation of the cylinder projecting the helix on to the Oxz plane, and the second the equation of the cylinder projecting it on to the Oyz plane. The cylinder projecting on to the Oxy plane is the cylinder $x^2 + y^2 = a^2$ on which the helix is actually traced. The projections of the helix on the co-ordinate planes Oxy, Oxz, Oyz may be seen to be respectively a circle, cosine and sine curves.

When the angle φ varies by 2π the point M rotates about the cylinder and at the same time rises by a height h equal to $2\pi c$. This quantity is called the pitch of the helix. A convenient form of the parametric equations of the helix is obtained by introducing h:

$$x = a\cos\varphi, \quad y = a\sin\varphi, \quad z = \frac{h}{2\pi}\varphi.$$

Both the present parameters (a and h) have a simple geometric meaning.

The helix has a number of interesting properties, two of which may be mentioned.

(1) We write down the equation for the tangent line to the helix at the point $M_0(x_0, y_0, z_0)$:

$$\frac{x - x_0}{-a\sin\varphi_0} = \frac{y - y_0}{a\cos\varphi_0} = \frac{z - z_0}{\frac{h}{2\pi}},$$

where φ_0 is the angle corresponding to M_0. Hence we have

$$\cos\gamma = \frac{h}{2\pi\sqrt{a^2 + \left(\frac{h}{2\pi}\right)^2}} = \frac{h}{\sqrt{(2\pi a)^2 + h^2}};$$

the direction cosine of the tangent with respect to Oz, i.e. the angle γ formed by the tangent with Oz, thus remains constant at every point of the helix. But the generators of the cylinder are parallel to Oz, so that the helix cuts the generators at a *constant angle*, which depends only on the radius of the cylinder and the pitch of the screw.

If we regard the generators as "meridians", the helix on a cylinder is a **loxodrome** (see Sec. 56).

(2) We unroll about a generator the cylinder on which the helix is traced so as to obtain a plane and confine ourselves to the segment of height equal to the pitch (this contains one turn of the helix) (Fig. 26). The base of the

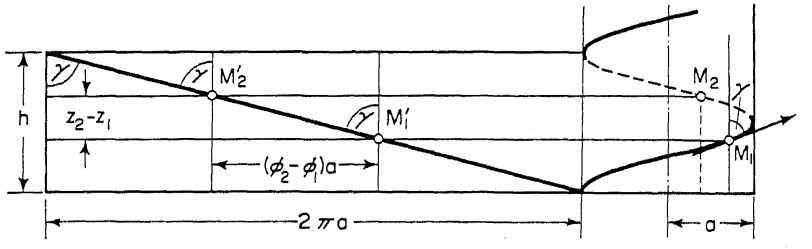

FIG. 26

cylinder and sections parallel to the base become parallel straight lines of length $2\pi a$, perpendicular to the straight lines of length h into which the generators are transformed. This unrolling does not distort the angles between curves traced on the cylinder nor the lengths of curves. In view of this the helix must unroll into a curve which cuts parallel straight lines in the same plane and at the same angle. But the only curve of this type is the straight line. Thus given the conditions of our construction the helix becomes the diagonal of a rectangle with sides $2\pi a$ and h. Fig. 26 clearly demonstrates the value found above for the cosine of the angle γ at which the helix cuts a generator.

Let M_1 and M_2 be any two points of the helix. The distance between them, measured along the helix, is equal to the straight segment M_1M_2 into which the arc M_1M_2 of the helix is transformed. We conclude from this that *the helix gives the shortest distance between two points of a cylinder.*

A curve on a surface passing through two given points and giving the shortest distance between them is called a **geodesic.** Thus the geodesic on a plane is a straight line, and on a sphere a great circle. The helix on a cylinder is not only a loxodrome but also a geodesic.

Let $\varphi = \varphi_1$, $z = z_1$ for the point M_1, $\varphi = \varphi_2$, $z = z_2$ for M_2. We easily find from Fig. 26 that the distance between these points measured along the helix is

$$\sqrt{(z_2 - z_1)^2 + (\varphi_2 - \varphi_1)^2 a^2} = \sqrt{\left(\frac{z_2 - z_1}{\varphi_2 - \varphi_1}\right)^2 + a^2}\,(\varphi_2 - \varphi_1)$$

$$= \sqrt{\left(\frac{h}{2\pi}\right)^2 + a^2}\,(\varphi_2 - \varphi_1).$$

The same expression is obtained by evaluating directly the arc length of the helix:

$$\int_{\varphi_1}^{\varphi_2} \sqrt{(-a\sin\varphi)^2 + (a\cos\varphi)^2 + \left(\frac{h}{2\pi}\right)^2}\, d\varphi = \sqrt{\left(\frac{h}{2\pi}\right)^2 + a^2} \int_{\varphi_1}^{\varphi_2} d\varphi$$

$$= \sqrt{\left(\frac{h}{2\pi}\right)^2 + a^2}\,(\varphi_2 - \varphi_1).$$

Let s denote the variable arc length of the helix corresponding to variation of the polar angle from zero to φ; now $s = m\varphi$, where $m = \sqrt{(h/2\pi)^2 + a^2}$. On substituting s/m for φ in the above equations we get the natural equations of the helix:

$$x = a\cos\frac{s}{m}, \quad y = a\sin\frac{s}{m}, \quad z = \frac{h}{2\pi m}\, s.$$

165. Curvature and Torsion. Frenet's Trihedral and Formulae*. We shall now investigate spatial curves in more detail.

There are two entities in the theory of spatial curves corresponding to curvature in the theory of plane curves: the *curvature*, measuring the deviation of the curve at a point from a straight line, and the *torsion*, measuring the deviation of the curve at a point from the plane. Spatial curves are therefore also called *curves of double curvature*.

We shall discuss these two entities by making use of vector analysis and considering the natural equation of the spatial curve:

$$\mathbf{r}(s) = x(s)\mathbf{i} + y(s)\mathbf{j} + z(s)\mathbf{k}.$$

I. Principal normal and curvature. We already know (Sec. 160) that the vector

$$\boldsymbol{\tau}_1 = \frac{d\mathbf{r}}{ds} = x'(s)\mathbf{i} + y'(s)\mathbf{j} + z'(s)\mathbf{k}$$

is the unit tangent to the curve at point $M(x, y, z)$, directed towards increasing s. Vector $\boldsymbol{\tau}_1$ is called the unit tangent.

Since

$$\boldsymbol{\nu} = \frac{d\boldsymbol{\tau}_1}{ds} = \frac{d^2\mathbf{r}}{ds^2} = x''(s)\mathbf{i} + y''(s)\mathbf{j} + z''(s)\mathbf{k}$$

is the derivative of the unit vector $\boldsymbol{\tau}_1$ it is perpendicular to $\boldsymbol{\tau}_1$, i.e. is normal to the curve at point M.

The straight line along which $\boldsymbol{\nu}$ is directed is called the principal normal to the curve.

If we take a unit vector $\boldsymbol{\nu}_1$, called the *unit principal normal*, in the direction of $\boldsymbol{\nu}$, we can write

$$\boldsymbol{\nu} = |\boldsymbol{\nu}|\,\boldsymbol{\nu}_1;$$

whilst the expression for the absolute value $|\boldsymbol{\nu}|$ is transformed as follows:

$$|\boldsymbol{\nu}| = \left|\frac{d\boldsymbol{\tau}_1}{ds}\right| = \left|\frac{d\boldsymbol{\tau}_1}{ds'}\right| \left|\frac{ds'}{ds}\right|,$$

* J. Frenet (1816–1900).

where s' is the length of the *hodograph* of vector $\boldsymbol{\tau}_1$. Hence $|d\boldsymbol{\tau}_1/ds'| = 1$, and since $\boldsymbol{\tau}_1$ is the unit tangent, ds' is equal to the *infinitesimal angle of displacement* $d\varphi$ of the curve at point M.

Thus

$$|\boldsymbol{\nu}| = \left|\frac{d\varphi}{ds}\right| \quad \text{and} \quad \boldsymbol{\nu} = \left|\frac{d\varphi}{ds}\right| \boldsymbol{\nu}_1.$$

By analogy with the plane curve (see Sec. 80) the absolute value of the rate of change of the tangential direction with the *length of arc*, i.e. $|d\varphi/ds|$, is taken as a measure of the bending of the curve at the point and is called the curvature. We denote the curvature by K:

$$K = \left|\frac{d\boldsymbol{\tau}_1}{ds}\right| = \left|\frac{d\varphi}{ds}\right| = |\boldsymbol{\nu}|.$$

We have in terms of the co-ordinates:

$$K = \sqrt{x''^2(s) + y''^2(s) + z''^2(s)}\,.$$

The curvature characterizes the degree of deviation of the curve at a point from the straight direction, i.e. from the direction of the tangent.

The reciprocal R of the curvature: $R = 1/K$, is called the *radius of curvature* at point M. We have

$$\boldsymbol{\nu}_1 = R\frac{d^2\mathbf{r}}{ds^2} \quad \text{and} \quad \boldsymbol{\nu} = K\boldsymbol{\nu}_1.$$

The radius of curvature is conveniently represented geometrically by a piece of length R of the principal normal with initial point at the point of the curve.

II. Binormals, Torsion and Frenet's formulae. We associate with each point M of a curve the unit tangent $\boldsymbol{\tau}_1$, the unit principal normal $\boldsymbol{\nu}_1$, and the unit vector $\boldsymbol{\beta}_1$ equal to the vector product of these unit vectors:

$$\boldsymbol{\beta}_1 = \boldsymbol{\tau}_1 \times \boldsymbol{\nu}_1.$$

We call $\boldsymbol{\beta}_1$ the unit binormal of the curve at M, the straight line on which it lies being the binormal. The binormal is perpendicular to vectors $\boldsymbol{\tau}_1$ and $\boldsymbol{\nu}_1$, i.e. it lies in the normal plane to the curve at M: it is therefore normal to the curve, like the "principal normal".

Vectors $\boldsymbol{\tau}_1, \boldsymbol{\nu}_1, \boldsymbol{\beta}_1$ form a system of vectors arranged like the fundamental unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$; they establish three main directions on a spatial curve. In the right-handed co-ordinate system that we always use, $\boldsymbol{\tau}_1, \boldsymbol{\nu}_1, \boldsymbol{\beta}_1$ form the vector triad shown in Fig. 27. Three mutually perpendicular planes pass through them and form a trihedral which is known as Frenet's trihedral. As the point M moves along the curve the Frenet trihedral changes its position in space and at each instant indicates the basic differential characteristics of the curve.

The plane through the tangent and principal normal (through $\boldsymbol{\tau}_1$ and $\boldsymbol{\nu}_1$) is called the *osculating plane**; the plane through the tangent and binormal

* The osculating plane at point M can be alternatively defined as the "limiting position" of the plane through M and two further points of the curve when these latter tend in an arbitrary manner to point M. We shall not prove this.

(through τ_1 and β_1) is the *rectifying plane;* and the plane through the principal normal and binormal (through ν_1 and β_1) is, as we already know, the *normal plane.*

Vector $d\beta_1/ds$, being the derivative of unit vector β_1, is perpendicular to β_1, i.e. lies in the osculating plane.

The equation

$$\beta_1 = \tau_1 \times \nu_1$$

gives us

$$\frac{d\beta_1}{ds} = \tau_1 \times \frac{d\nu_1}{ds} + \frac{d\tau_1}{ds} \times \nu_1,$$

and since $d\tau_1/ds = \nu$, the second term is the vector product of collinear vectors and therefore vanishes: thus

$$\frac{d\beta_1}{ds} = \tau_1 \times \frac{d\nu_1}{ds}$$

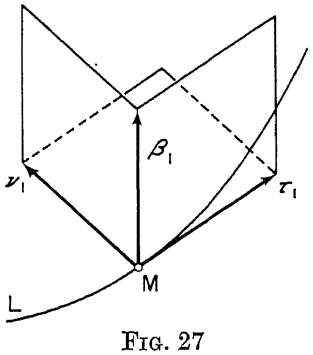

FIG. 27

and $d\beta_1/ds$ is perpendicular to τ_1. Since $d\beta_1/ds$ is both perpendicular to the tangent and lies in the osculating plane, it follows that it is directed along the *principal normal.* The geometrical significance of this vector may be seen with the aid of the same arguments as were applied to vector $d\tau_1/ds$, viz

$$\frac{d\beta_1}{ds} = \frac{d\beta_1}{ds''}\frac{ds''}{ds},$$

where s'' is the arc length of the hodograph of vector β_1. Hence $|d\beta_1/ds''| = 1$, and since β_1 is the unit binormal vector, ds'' is equal to the *infinitesimal angle $d\psi$ of rotation of the binormal.*

The quantity $|d\psi/ds|$ gives the rate of change of the direction of the binormal (or of the osculating plane) with respect to the arc of the curve; when it is supplied with the + or — sign, it is taken as a measure of the "twist" of curve L at the point in question and is called the *torsion.* We denote it by T:

$$T = \pm\left|\frac{d\beta_1}{ds}\right| = \pm\left|\frac{d\psi}{ds}\right|.$$

It has the + sign if the direction of vector $d\beta_1/ds$ coincides with that of ν_1, and the — sign in the opposite case.

The torsion *characterizes the degree of deviation of the curve at a point from a plane, i.e. from the corresponding osculating plane.*

In the case of a plane curve all the tangents lie in one plane and the principal normal lies in the same plane, the principal normal being now simply the normal; whilst the osculating plane coincides with the plane of the curve. *The torsion of a plane curve is zero.* The greater the torsion, the greater the deviation of the curve from a plane at a given point.

The reciprocal R_1 of the torsion: $R_1 = 1/T$, is called the *radius of torsion* of the curve at point M.

Bearing in mind that vector $d\beta_1/ds$ is directed along or in opposition to the unit principal normal, we can write

$$\frac{d\beta_1}{ds} = \pm\left|\frac{d\beta_1}{ds}\right|\nu_1 = T\nu_1.$$

We have already found the rates of change of unit vectors $\boldsymbol{\tau}_1$ and $\boldsymbol{\beta}_1$ with respect to the arc length s; it remains for us to find $d\boldsymbol{\nu}_1/ds$—the rate of change of unit vector $\boldsymbol{\nu}_1$.

We have:

$$\boldsymbol{\nu}_1 = -(\boldsymbol{\tau}_1 \times \boldsymbol{\beta}_1),$$

whence

$$\frac{d\boldsymbol{\nu}_1}{ds} = -\frac{d(\boldsymbol{\tau}_1 \times \boldsymbol{\beta}_1)}{ds} = -\left(\boldsymbol{\tau}_1 \times \frac{d\boldsymbol{\beta}_1}{ds} + \frac{d\boldsymbol{\tau}_1}{ds} \times \boldsymbol{\beta}_1\right)$$

$$= -(\boldsymbol{\tau}_1 \times T\boldsymbol{\nu}_1 + K\boldsymbol{\nu}_1 \times \boldsymbol{\beta}_1) = -(T\boldsymbol{\beta}_1 + K\boldsymbol{\tau}_1 = -\frac{\boldsymbol{\tau}_1}{R} - \frac{\boldsymbol{\beta}_1}{R_1}.$$

The three formulae that we have obtained:

$$\frac{d\boldsymbol{\tau}_1}{ds} = \frac{\boldsymbol{\nu}_1}{R},$$

$$\frac{d\boldsymbol{\nu}_1}{ds} = -\frac{\boldsymbol{\tau}_1}{R} - \frac{\boldsymbol{\beta}_1}{R_1},$$

$$\frac{d\boldsymbol{\beta}_1}{ds} = \frac{\boldsymbol{\nu}_1}{R_1}$$

are called *Frenet's formulae*. They play an important part in the theory of spatial curves.

Frenet's formulae connect the rates of change of the basic unit vectors of the curve with its radii of curvature and torsion. On projecting each of these equations on to the co-ordinate axes their co-ordinate forms can be obtained, though we shall not dwell on this. We shall merely find the co-ordinate expression for the torsion T.

We shall start from the second Frenet formula:

$$T\boldsymbol{\beta}_1 = -K\boldsymbol{\tau}_1 - \frac{d\boldsymbol{\nu}_1}{ds};$$

since $K\boldsymbol{\nu}_1 = \boldsymbol{\nu}$ and $KR = 1$, we can rewrite this as

$$T\boldsymbol{\beta}_1 = -K\boldsymbol{\tau}_1 - R\frac{d\boldsymbol{\nu}}{ds} - \boldsymbol{\nu}\frac{dR}{ds}.$$

On forming the scalar product of this equation with $\boldsymbol{\beta}_1$ and recalling that $\boldsymbol{\beta}_1 \cdot \boldsymbol{\beta}_1 = 1$, $\boldsymbol{\beta}_1 \cdot \boldsymbol{\tau}_1 = \boldsymbol{\beta}_1 \cdot \boldsymbol{\nu} = 0$, we get

$$T = -R\left(\boldsymbol{\beta}_1 \cdot \frac{d\boldsymbol{\nu}}{ds}\right).$$

Bearing in mind that

$$\boldsymbol{\beta}_1 = \boldsymbol{\tau}_1 \times \boldsymbol{\nu}_1 = (x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}) \times (Rx''\mathbf{i} + Ry''\mathbf{j} + Rz''\mathbf{k})$$

and

$$\frac{d\boldsymbol{\nu}}{ds} = x'''\mathbf{i} + y'''\mathbf{j}$$

we now get

$$T = -R\left(\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x' & y' & z' \\ Rx'' & Ry'' & Rz'' \end{vmatrix} \cdot (x'''\mathbf{i} + y'''\mathbf{j} + z'''\mathbf{k})\right).$$

On taking R outside the brackets, replacing $\mathbf{i}, \mathbf{j}, \mathbf{k}$ in the determinant by x''', y''', z''' respectively and rearranging the rows, we obtain the final expression for T as a third order determinant:

$$\mathrm{T} = -R^2 \begin{vmatrix} x'(s) & y'(s) & z'(s) \\ x''(s) & y''(s) & z''(s) \\ x'''(s) & y'''(s) & z'''(s) \end{vmatrix}$$

where

$$R = \frac{1}{\sqrt{x''^2(s) + y''^2(s) + z''^2(s)}}.$$

III. Elements of the trihedral. We shall end by finding the equations of the three ribs (i.e. *the tangent, principal normal and binormal*) and the equations of the three faces (i.e. *the osculating, rectifying and normal planes*) of the trihedral of the curve $\mathbf{r} = x(s)\mathbf{i} + y(s)\mathbf{j} + z(s)\mathbf{k}$ at a point $M_0(x_0, y_0 z_0)$, where $x_0 = x(s_0)$, $y_0 = y(s_0)$, $z_0 = z(s_0)$. We put

$$x_0' = x'(s_0), \quad y_0' = y'(s_0), \quad z_0' = z'(s_0);$$

$$x_0'' = x''(s_0), \quad y_0'' = y''(s_0), \quad z_0'' = z''(s_0).$$

I′. *The tangent* is

$$\frac{x - x_0}{x_0'} = \frac{y - y_0}{y_0'} = \frac{z - z_0}{z_0'},$$

since the projections of $\boldsymbol{\tau}_1$ are x_0', y_0', z_0'.

II′. The *principal normal* is

$$\frac{x - x_0}{x_0''} = \frac{y - y_0}{y_0''} = \frac{z - z_0}{z_0''},$$

since the projections of $\boldsymbol{\nu}_1$ are Rx_0'', Ry_0'', Rz_0''.

III′. The *binormal* is

$$\frac{x - x_0}{y_0' z_0'' - y_0'' z_0'} = \frac{y - y_0}{z_0' x_0'' - z_0'' x_0'} = \frac{z - z_0}{x_0' y_0'' - x_0'' y_0'},$$

since the projections of $\boldsymbol{\beta}_1$ are

$$R[y_0' z_0'' - y_0'' z_0'], \quad R[z_0' x_0'' - z_0'' x_0'], \quad R[x_0' y_0'' - x_0'' y_0'].$$

I″. The *osculating plane* is

$$(y_0' z_0'' - y_0'' z_0')(x - x_0) + (z_0' x_0'' - z_0'' x_0')(y - y_0) + \\ + (x_0' y_0'' - x_0'' y_0')(z - z_0) = 0,$$

since it passes through point M_0 and is perpendicular to the binormal.

II″. The *rectifying plane* is

$$x_0''(x - x_0) + y_0''(y - y_0) + z_0''(z - z_0) = 0,$$

since it passes through M_0 and is perpendicular to the principal normal.

III″. The *normal plane* is

$$x_0'(x - x_0) + y_0'(y - y_0) + z_0'(z - z_0) = 0,$$

since it passes through M_0 and is perpendicular to the tangent.

Example. We shall take the helix for illustration (Sec. 164):

$$\mathbf{r} = a\cos\frac{s}{m}\,\mathbf{i} + a\sin\frac{s}{m}\,\mathbf{j} + \frac{h}{2\pi m}\,s\mathbf{k}.$$

The tangent to the helix and associated properties were found in Sec. 164. The equations of the principal normal at $M_0(x_0, y_0, z_0)$ are

$$\frac{x - x_0}{-\frac{a}{m^2}\cos\frac{s_0}{m}} = \frac{y - y_0}{-\frac{a}{m^2}\sin\frac{s_0}{m}} = \frac{z - z_0}{0},$$

i.e.

$$\frac{x - x_0}{x_0} = \frac{y - y_0}{y_0} \quad \text{and} \quad z = z_0,$$

or finally,

$$\frac{x}{x_0} = \frac{y}{y_0}, \quad z = z_0.$$

Thus the *principal normal at any point of the helix intersects and is perpendicular to the axis of the cylinder.*

Similarly, we find for the equations of the binormal:

$$\frac{x - x_0}{y_0} = \frac{y - y_0}{-x_0} = \frac{z - z_0}{\frac{2\pi}{h}a^2}.$$

This shows that the *binormal at any point of the helix forms a constant angle with the axis of the cylinder, the cosine of which is a/m.*

We have for the radius of curvature R:

$$R = \frac{1}{\sqrt{\frac{a^2}{m^4}\cos^2\frac{s_0}{m} + \frac{a^2}{m^4}\sin^2\frac{s_0}{m} + 0}} = \frac{m^2}{a},$$

i.e. *the radius of curvature is constant.*

We have for the radius of torsion R_1:

$$R_1 = -\frac{1}{R^2\begin{vmatrix} -\frac{a}{m}\sin\frac{s_0}{m} & \frac{a}{m}\cos\frac{s_0}{m} & \frac{h}{2\pi m} \\ -\frac{a}{m^2}\cos\frac{s_0}{m} & -\frac{a}{m^2}\sin\frac{s_0}{m} & 0 \\ \frac{a}{m^3}\sin\frac{s_0}{m} & -\frac{a}{m^3}\cos\frac{s_0}{m} & 0 \end{vmatrix}}$$

$$= -\frac{1}{\frac{m^4}{a^2}\,\frac{h}{2\pi m}\,\frac{a^2}{m^5}} = -\frac{2\pi}{h}m^2,$$

i.e. *the radius of torsion is also constant.* The helix is the only spatial curve having both radii constant.

REMARK. If the curve is given by the equation $\boldsymbol{r} = x(t)\,\boldsymbol{i} + y(t)\,\boldsymbol{j} + z(t)\,\boldsymbol{k}$ (or what amounts to the same thing, by the three equations $x = x(t)$, $y = y(t)$, $z = z(t)$), in which parameter t is not the arc length s, the equations of the principal normal (II′) and of the rectifying plane (II″) must be changed as follows: instead of x_0'', y_0'', z_0'' we have to take respectively $s_0'x_0'' - s_0''x_0'$, $s_0'y_0'' - s_0''y_0'$, $s_0'z_0'' - s_0''z_0'$, where s_0 is the arc length corresponding to point M_0, and

$$s = \int_0^t \sqrt{x'^2(t) + y'^2(t) + z'^2(t)}\,dt = \varphi(t);$$

here $x(t_0) = x_0$, $x'(t_0) = x_0'$, $x''(t_0) = x_0''$ and so on, and $s_0 = \varphi(t_0)$. The equations of the tangent (I′), binormal (III′), osculating plane (I″) and normal plane (III″) remain unchanged. We have, in fact:

$$\frac{dx}{ds} = x'(t)\frac{dt}{ds} = \frac{x'(t)}{\varphi'(t)},$$

$$\frac{dy}{ds} = y'(t)\frac{dt}{ds} = \frac{y'(t)}{\varphi'(t)},$$

$$\frac{dz}{ds} = z'(t)\frac{dt}{ds} = \frac{z'(t)}{\varphi'(t)};$$

hence x_0', y_0', z_0' (differentiation with respect to s) are proportional to $x'(t_0)$, $y'(t_0)$, $z'(t_0)$; furthermore,

$$\frac{d^2x}{ds^2} = \frac{x''(t)\varphi'(t) - \varphi''(t)x'(t)}{\varphi'^3(t)},$$

$$\frac{d^2y}{ds^2} = \frac{y''(t)\varphi'(t) - \varphi''(t)y'(t)}{\varphi'^3(t)},$$

$$\frac{d^2z}{ds^2} = \frac{z''(t)\varphi'(t) - \varphi''(t)z'(t)}{\varphi'^3(t)};$$

hence x_0'', y_0'', z_0'' (differentiation with respect to s) are proportional to $x''(t_0)\varphi'(t_0) - \varphi''(t_0)x'(t_0)$, $y''(t_0)\varphi'(t_0) - \varphi''(t_0)y'(t_0)$, $z''(t_0)\varphi'(t_0) - \varphi''(t_0)z'(t_0)$; finally,

$$\frac{dy}{ds}\frac{d^2z}{ds^2} - \frac{d^2y}{ds^2}\frac{dz}{ds} = \frac{y'(t)z''(t) - y''(t)z'(t)}{\varphi'^3(t)},$$

$$\frac{dz}{ds}\frac{d^2x}{ds^2} - \frac{d^2z}{ds^2}\frac{dx}{ds} = \frac{z'(t)x''(t) - z''(t)x'(t)}{\varphi'^3(t)},$$

$$\frac{dx}{ds}\frac{d^2y}{ds^2} - \frac{d^2x}{ds^2}\frac{dy}{ds} = \frac{x'(t)y''(t) - x''(t)y'(t)}{\varphi'^3(t)};$$

hence $y_0'z_0'' - y_0''z_0'$, $z_0'x_0'' - z_0''x_0'$, $x_0'y_0'' - x_0''y_0'$ (differentiation with respect to s) are proportional to $y'(t_0)z''(t_0) - y''(t_0)z'(t_0)$, $z'(t_0)x''(t_0) - z''(t_0)x'(t_0)$, $x'(t_0)y''(t_0) - x''(t_0)y'(t_0)$.

This is what we required to prove.

166. Surfaces. In Sec. 146 we defined the tangent plane to the surface $z = f(x, y)$ as the plane through the tangent lines to two plane sections of the surface parallel to the co-ordinate planes Oxz and Oyz. We shall now dwell in more detail on the properties of tangent planes.

Let the surface be given by the equation

$$F(x, y\, z) = 0.$$

We shall suppose that function F is differentiable for the relevant values of the arguments; also, that not all the partial deri-

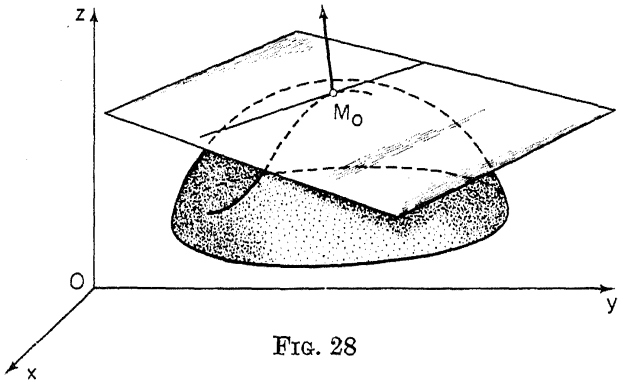

FIG. 28

vatives $\partial F/\partial x$, $\partial F/\partial y$, $\partial F/\partial z$ vanish at a point $M_0(x_0, y_0, z_0)$ of the surface, i.e. that M_0 is not a singular point (the definition of singular point for the surface $F(x, y, z) = 0$ is similar to the definition for the curve $F(x, y) = 0$, see Sec. 162).

We draw any curve on the surface (not necessarily a plane curve) through M_0 which has a tangent line at M_0 (Fig. 28). This line is also a tangent line to the surface.

THEOREM. **All the tangent lines to a surface at point M_0 lie in the tangent plane at M_0.**

Proof. If

$$x = x(t), \quad y = y(t), \quad z = z(t),$$

are the equations of the curve on the surface, the equations of the tangent line at M_0 can be written as (Sec. 164):

$$\frac{x - x_0}{x'_0} = \frac{y - y_0}{y'_0} = \frac{z - z_0}{z'_0}, \qquad (*)$$

where $x_0 = x(t_0)$, $y_0 = y(t_0)$, $z_0 = z(t_0)$ and $x'_0 = x'(t_0)$, $y'_0 = y'(t_0)$, $z'_0 = z'(t_0)$. Since the curve lies on the surface, functions $x(t)$, $y(t)$, $z(t)$ must satisfy its equation $F(x, y, z) = 0$. Differentiation

of F with respect to t gives us, on the assumption that x, y, z are the functions of t in question:

$$\frac{\partial F}{\partial x}x' + \frac{\partial F}{\partial y}y' + \frac{\partial F}{\partial z}z' = 0.$$

This equation holds for all points of the curve, and in particular, for M_0:

$$\left(\frac{\partial F}{\partial x}\right)_0 x'_0 + \left(\frac{\partial F}{\partial y}\right)_0 y'_0 + \left(\frac{\partial F}{\partial z}\right)_0 z'_0 = 0. \qquad (**)$$

The zero subscript indicates that the partial derivatives are taken at M_0. We now write the equation of the straight line through M_0 with direction cosines proportional to $(\partial F/\partial x)_0$, $(\partial F/\partial y)_0$, $(\partial F/\partial z)_0$:

$$\frac{x - x_0}{\left(\frac{\partial F}{\partial x}\right)_0} = \frac{y - y_0}{\left(\frac{\partial F}{\partial y}\right)_0} = \frac{z - z_0}{\left(\frac{\partial F}{\partial z}\right)_0}. \qquad (***)$$

It is well known from analytic geometry that equation (**) expresses the perpendicularity of straight lines (*) and (***). But the latter, as is clear from its equation, depends only on the surface and its point M_0, and not at all on the curve taken on the surface. Consequently any tangent line to the surface at M_0 is perpendicular to the same straight line (***), i.e. all the tangent lines lie in a plane perpendicular to (***). It remains to show that this plane is in fact the tangent plane to the surface.

On again starting from the familiar geometrical condition for a straight line and plane to be perpendicular, we can form the equation of the plane through M_0 perpendicular to the straight line (***):

$$\left(\frac{\partial F}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial F}{\partial y}\right)_0 (y - y_0) + \left(\frac{\partial F}{\partial z}\right)_0 (z - z_0) = 0;$$

this is in fact the *equation of the tangent plane to the surface* $F(x, y, z) = 0$ *at the point* $M_0(x_0, y_0, z_0)$. For, the equation can be rewritten as (provided $(\partial F/\partial z)_0 \neq 0$):

$$z - z_0 = -\frac{\left(\frac{\partial F}{\partial x}\right)_0}{\left(\frac{\partial F}{\partial z}\right)_0}(x - x_0) - \frac{\left(\frac{\partial F}{\partial y}\right)_0}{\left(\frac{\partial F}{\partial z}\right)_0}(y - y_0),$$

or (see Sec. 151):

$$z - z_0 = \left(\frac{\partial z}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial z}{\partial y}\right)_0 (y - y_0).$$

We have arrived at the equation of the tangent plane obtained in Sec. 146.

This is what we had to prove.

Definition. **The straight line (***), perpendicular to the tangent plane at the point of contact, is called the *normal to the surface* at this point** (see Sec. 161).

If the equation of the surface is given in the form $z = f(x, y)$, the equation of the normal at point $M_0(x_0, y_0, z_0)$ can be written as

$$\frac{x - x_0}{-f'_x(x_0, y_0)} = \frac{y - y_0}{-f'_y(x_0, y_0)} = \frac{z - z_0}{1}.$$

We have for the direction cosines $\cos\alpha$, $\cos\beta$, $\cos\gamma$ of the normal:

$$\cos\alpha = \frac{-f'_x(x_0, y_0)}{\sqrt{1 + f'^2_x(x_0, y_0) + f'^2_y(x_0, y_0)}},$$

$$\cos\beta = \frac{-f'_y(x_0, y_0)}{\sqrt{1 + f'^2_x(x_0, y_0) + f'^2_y(x_0, y_0)}},$$

$$\cos\gamma = \frac{1}{\sqrt{1 + f'^2_x(x_0, y_0) + f'^2_y(x_0, y_0)}}.$$

Example. Let us find the tangent plane and normal to the sphere

$$x^2 + y^2 + z^2 = R$$

at the point $M_0(x_0, y_0, z_0)$.

We have:

$$\left(\frac{\partial F}{\partial x}\right)_0 = 2x_0, \quad \left(\frac{\partial F}{\partial y}\right)_0 = 2y_0, \quad \left(\frac{\partial F}{\partial z}\right)_0 = 2z_0,$$

i.e. the equation of the tangent plane is

$$x_0(x - x_0) + y_0(y - y_0) + z_0(z - z_0) = 0$$

or

$$x_0 x + y_0 y + z_0 z = x_0^2 + y_0^2 + z_0^2 = R^2.$$

The equations of the normal are

$$\frac{x - x_0}{x_0} = \frac{y - y_0}{y_0} = \frac{z - z_0}{z_0} \quad \text{or} \quad \frac{x}{x_0} = \frac{y}{y_0} = \frac{z}{z_0}.$$

It follows from this that the normal to the sphere passes through the origin, i.e that the normals are radii.

CHAPTER XII

MULTIPLE INTEGRALS AND ITERATED INTEGRATION

1. Double and Triple Integrals

167. Problems on Volumes. Double Integrals. We now turn to a discussion of the integral calculus for functions of several independent variables. We shall arrive at the concept of a double (then of a triple) integral by starting as previously from a concrete geometrical problem—in this case the problem of finding the volume of a solid.

Let us formulate the problem of finding the volume of a given solid bounded by a certain surface. We refer the surface to a system of Cartesian co-ordinates $Oxyz$ in space and start by considering what we shall term a "cylindrical solid."

Definition. A cylindrical solid *is one which is bounded by the Oxy plane, by a surface which is cut in not more than one point by any straight line parallel to Oz, and by a cylindrical surface whose generators are parallel to Oz.*

The domain D cut out of the Oxy plane by the cylindrical surface is called the base *of the cylindrical solid* (Fig. 29).

The base of a cylindrical solid is the orthogonal projection of the bounding surface of the solid on the Oxy plane.

A solid can usually be made up from cylindrical solids and the required volume defined as the sum of the volumes of the component cylindrical solids. It follows from this that we only need to find the volume of a cylindrical solid in order to solve the present problem.

We shall first of all recall two principles which have already been mentioned (Sec. 119) in regard to finding the volume of a solid.

(1) *If a solid is divided into parts, its volume is equal to the sum of the volumes of all the parts* (property of additiveness).

(2) *The volume of a cylindrical body bounded by a plane parallel to the Oxy plane is equal to the base area times the height.*

Let $z = f(x, y)$ be the equation of the surface bounding the cylindrical body. We shall assume that $f(x, y)$ is a continuous function of the point $P(x, y)$ in domain D and, as a preliminary, that the surface lies wholly above the Oxy plane, i.e. that $f(x, y) > 0$ everywhere in domain D.

Let V denote the required volume of the cylindrical body.

We divide the base — the domain D — into a number n of non-intersecting domains of arbitrary shape; we shall call these sub-domains. We enumerate the sub-domains in a certain order and

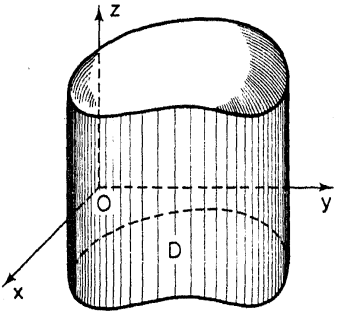

FIG. 29

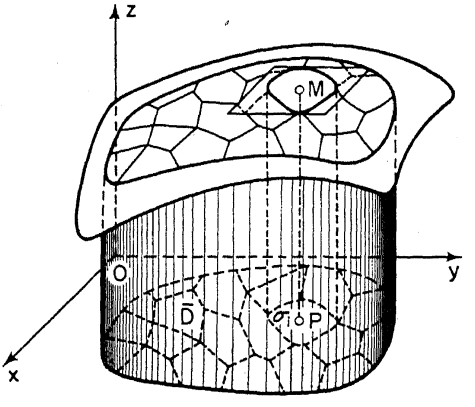

FIG. 30

denote them by $\sigma_1, \sigma_2, \ldots, \sigma_n$, their areas being $\Delta\sigma_1, \Delta\sigma_2, \ldots \Delta\sigma_n$. We draw a cylindrical surface (with generators parallel to Oz) through the boundary of each sub-domain. These cylindrical surfaces divide the surface into n pieces, corresponding to the n sub-domains. The cylindrical body has thus been divided into n partial cylindrical bodies (Fig. 30). This sub-division tells us nothing in itself, since the precise definition of the volume of a partial cylindrical body is just as difficult as the original problem. The real point is that the bases have diminished with sub-division. The bases of the sub-domains are required to tend to zero in what follows. *The diameter of a finite domain is defined as the greatest distance between two points of its boundary.* If the diameter tends to zero, the domain shrinks to a point.

Further, we replace the surface bounding the i-th partial cylinder by a piece of plane parallel to Oxy and distant from Oxy by an amount equal to the z co-ordinate of any point whatever of the original surface. We obtain as a result an n-step solid the volume V_n of which is readily determined.

Let us find the volume of the i-th cylinder. Its height is equal to a z co-ordinate of the surface, i.e. the value of $z = f(x, y)$ at some point of the domain σ_i; we write $P_i(\xi_i, \eta_i)$ for this point. Thus the volume of the i-th cylinder is

$$f(\xi_i, \eta_i)\,\Delta\sigma_i \quad (= f(P_i)\,\Delta\sigma_i).$$

Consequently

$$V_n = f(\xi_1, \eta_1)\,\Delta\sigma_1 + f(\xi_2, \eta_2)\,\Delta\sigma_2 + \cdots + f(\xi_n, \eta_n)\,\Delta\sigma_n$$

$$= \sum_{i=1}^{n} f(\xi_i, \eta_i)\,\Delta\sigma_i$$

or, more briefly,

$$V_n = \sum_{i=1}^{n} f(P_i)\,\Delta\sigma_i. \qquad (*)$$

We take the volume V of the original cylindrical body as approximately equal to the volume of the n-step solid, with the assumption that V_n expresses V the more accurately, the smaller the *greatest* of the diameters of the sub-domains, i.e. the greater n.

By virtue of this, *we take the required volume V as equal by definition to the limit of sum* (*) *as* $n \to \infty$ *and the greatest of the sub-domain diameters tends to zero.*

Sum (*) *is called the* n-th integral sum *for function* $f(x, y)$ *in domain* D, *corresponding to the sub-division of* D *into* n *sub-domains.*

Definition. **The limit to which the n-th integral sum (*) tends as the greatest of the sub-domain diameters tends to zero is called the *double integral of function $f(x, y)$ over domain D.***

We write this as:

$$\lim \sum_{i=1}^{n} f(P_i)\,\Delta\sigma_i = \iint_D f(P)\,d\sigma = \iint_D f(x, y)\,d\sigma.$$

The double integral could be denoted by the single symbol $\int$, but two such symbols $\iint$ are generally used for convenience in later working.

This is read as: "the double integral over domain D of $f(x, y)\ d\sigma$." The symbol $d\sigma$ indicates the indefinitely diminishing area of a sub-domain *(differential* or *element of area)*; $f(P)\ d\sigma$, indicating the form of the terms to be summed, is called the *integrand element;* $f(P)$ is called the *integrand;* the letter D below the double integral sign shows the domain of the Oxy plane over which the summation has been carried out; finally, variables x and y and the point $P(x, y)$

are called respectively the *variables of integration* and the *variable point of integration.*

We can therefore say that *the volume of a cylindrical body bounded by the Oxy plane, the surface* $z = f(x, y)$ $(f(x, y) > 0)$ *and a cylindrical surface with generators parallel to Oz is given by the double integral of the z co-ordinate of the surface, i.e. of function* $f(x, y)$, *over the domain consisting of the base of the cylindrical body:*

$$V = \iint_D f(x, y)\, d\sigma .$$

A wide variety of problems, apart from those on volumes, leads to the formation of the sum (*) for functions of two independent variables and to subsequent passage to the limit. We must therefore extend our definition of double integral to *any* continuous function $z = f(x, y)$ in domain D, independently of the actual physical nature of the variables x and y and of their function $f(x, y)$, and without the restriction that $f(x, y) > 0$.

Some remarks need to be made in regard to the present definition of double integral analogous to those made when defining the ordinary integral (Sec. 86). These remarks lead us to the following theorem.

EXISTENCE THEOREM FOR DOUBLE INTEGRALS*. **The n-th integral sum corresponding to a finite domain D of variation of point $P(x, y)$ and to a function $f(P)$ continuous in this domain tends to a limit as $n \to \infty$ and the greatest sub-domain diameter tends to zero. This limit is independent of the manner of sub-division of D into sub-domains and of the points P_i chosen in the sub-domains. It is called the *double integral of function* $f(P)$ *over the domain* D.**

It should be noted that the domain of integration can be either singly or multiply connected (see Sec. 139).

The double integral is, of course, a number which depends only on the integrand and the domain of integration, and not at all on the notation for the variables of integration, so that, for example,

$$\iint_D f(x, y)\, d\sigma = \iint_D f(u, v)\, d\sigma .$$

We shall see later (Art. 2) that a double integral can be evaluated by means of ordinary integrations.

* For the proof, see e.g. G.M. FIKHTENGOL'TS, *Course of Differential and Integral Calculus* (*Kurs differentsial'nogo i integral'nogo ischisleniya*), vol. III, pp. 150–169, Gost., 1949; R. COURANT, *Course of Differential and Integral Calculus*, Part II, p. 231 *et seq.*, 1931.

168. General Definition of Integral. Triple Integrals. The constructions for ordinary and double integrals are entirely analogous. The analogy becomes even more obvious if we regard functions of either one or two independent variables as functions of a point. The integral sum for a function of one independent variable $f(P) = f(x)$ may be written as

$$l_n = \sum_{i=1}^{n} f(P_i)\, \Delta l_i,$$

where Δl_i is the length of the i-th sub-interval, and P_i is an arbitrary point in the sub-interval. It must be observed, however, that Δl_i is a number possessing a sign, i.e. that the "directed" length of the sub-interval is taken, and not simply the geometric measure of the sub-domain, as in the case of the double integral. Here, however, we shall pay no attention to the sign and shall regard Δl simply as an element of length. The integral of $f(P)$ in a given interval L can be denoted by the symbol

$$l = \int_L f(P)\, dl,$$

where dl is the element (differential) of *length* in interval L. Interval L is also termed the *domain of integration.*

Let us compare the expressions for the integral sum and integral in the case of a function of one variable with the corresponding expressions for a function of two variables. It will be seen that the difference between the integrals lies in the nature of the domain of variation of point P and hence in the nature of the summation, and not in the symbolism or structure of the formulae. In the first case the variable point of integration P moves along the axis of the independent variable, and an element of the integral is got by multiplying the value of the function by an *element of length* of the domain of integration (which is part of the real axis). The ordinary integral is therefore said to be single or rectilinear. In the second case the variable point of integration P moves over the plane of the independent variables, and an element of the integral is got by multiplying the value of the function by an *element of area* of the domain of integration (which is part of the plane). The integral in this case is therefore said to be *double.*

A common definition can therefore yield either the ordinary or the double integral.

Definition. ***The integral I of a function $f(P)$ in a finite domain W in which the function is continuous is defined as the limit of the n-th integral sum***

$$I_n = \sum_{i=1}^{n} f(P_i)\, \Delta w_i$$

as $n \to \infty$ and the greatest sub-domain diameter w_i tends to zero. The integral sum I_n is composed for an arbitrary subdivision of the domain W into n sub-domains, P_i is an arbitrary point of the i-th sub-domain, and Δw_i denotes the measure of the sub-domain.

This general definition of the ordinary and double integral can be extended to functions of any number of independent variables. The integral thus defined is said to be *multiple*. Let W be the domain Ω of space of three independent variables, in which the variable point of integration $P(x, y, z)$ moves; Δw_i is now the volume Δv_i of sub-domain v_i and dw is the differential dv of this volume. We obtain an integral which is generally written by means of three symbols $\int$:

$$I = \iiint_{\Omega} f(P)\, dv = \iiint_{\Omega} f(x, y, z)\, dv$$

and is described as *triple*.

An "existence theorem" for the triple integral can be formulated precisely as in the case of the double integral. The terminology used for ordinary and double integrals can be carried over to triple integrals.

We have arrived here at the concept of triple integral by a purely formal extension of the concepts of ordinary and double integrals, but we shall shortly encounter concrete problems, the solutions of which lead to triple integrals.

169. Fundamental Properties of Double and Triple Integrals. In view of their common definition, double, triple and in general multiple integrals possess properties similar to those of the ordinary integral. This will be explained by reference to the double integral.

THEOREM I (on the integral of a sum). **The double integral of the sum of a finite number of functions is equal to the sum of the double integrals of the individual functions:**

$$\iint_D [f(P) + \varphi(P) + \cdots + \psi(P)]\, d\sigma$$

$$= \iint_D f(P)\, d\sigma + \iint_D \varphi(P)\, d\sigma + \cdots + \iint_D \psi(P)\, d\sigma.$$

THEOREM II (on moving a constant factor outside). **A constant factor in the integrand can be taken outside the symbol of double integration:**

$$\iint_D cf(P)\,d\sigma = c\iint_D f(P)\,d\sigma.$$

These theorems are obtained directly by applying the familiar rules for passage to the limit in the corresponding integral sums.

THEOREM III (on the sign of the integral). **If the integrand does not change sign in the domain of integration, the double integral is a number having the same sign as the integrand.**

Proof. Let $f(P) \geqslant 0$ in domain D. All the terms are now non-negative in the integral sum

$$I_n = \sum_{i=1}^{n} f(P_i)\,\Delta\sigma_i$$

so that $I_n \geqslant 0$; and the limit of a non-negative quantity cannot be negative. The integral of a continuous function $f(P)$ of constant sign can only be equal to zero in the case when $f(P)$ is identically zero. This is proved in the same way as for a function of one independent variable (Sec. 90).

If the integrand changes sign in the domain of integration, its integral may be either positive or negative, or equal to zero.

THEOREM IV (on sub-division of the domain of integration; property of additiveness). **If the domain of integration D is divided into two parts D_1 and D_2, we have**

$$\iint_D f(P)\,d\sigma = \iint_{D_1} f(P)\,d\sigma + \iint_{D_2} f(P)\,d\sigma.$$

Proof. Since the limit of an integral sum does not depend on the method of sub-division of domain D, we can divide D in such a way that each sub-domain σ_i belongs either wholly in D_1 or wholly in D_2; the integral sum can now be written as

$$\sum f(P_i)\,\Delta\sigma_i = \sum_1 f(P_i)\,\Delta\sigma_i + \sum_2 f(P_i)\,\Delta\sigma_i,$$

where all the elements corresponding to sub-domains belonging to D_1 are collected in the first sum on the right, and all the elements corresponding to sub-domains belonging to D_2 are in the second sum. On passing to the limit on the assumption that the greatest sub-domain diameter for the whole of D tends to zero, we obtain the required equation.

It follows directly from this that, if D is divided into k partial domains $D_1, D_2, \ldots, D_k$, we have

$$\iint_D f(P)\,d\sigma = \iint_{D_1} f(P)\,d\sigma + \iint_{D_2} f(P)\,d\sigma + \cdots + \iint_{D_k} f(P)\,d\sigma.$$

We now turn the geometrical interpretation of the double integral. We agree to write the plus sign in front of the volume of a cylindrical body located above the Oxy plane, and minus if it is located below Oxy. It is now obvious that the double integral of a function $f(P)$, regarded as the z co-ordinate of some bounded surface, is the algebraic sum of the "volumes" of the cylindrical bodies corresponding to positive and negative values of $f(P)$.

Bearing this in mind, we can in future interpret the double integral

$$\iint_D f(x, y)\,d\sigma,$$

independently of the concrete meaning of the variables of integration x and y and of function $f(x, y)$, as the "volume" (algebraic, not geometric) of the cylindrical body with base D bounded by the surface $z = f(x, y)$.

On the other hand, if we want to find the true (geometric) volume of a cylindrical body, we have to evaluate separately the integral giving the volume of the part above the Oxy plane and the integral giving the part below Oxy, then take the sum of the absolute values of these integrals.

THEOREM V (on the upper and lower bounds of an integral). **The value of a double integral lies between the products of the greatest and least values of the integrand with the area of the domain of integration, i.e.**

$$\boldsymbol{mS \leqslant \iint_D f(P)\,d\sigma \leqslant MS,}$$

where m and M are respectively the least and greatest values of $f(P)$ in domain D and S is the area of D.

We consider functions $M - f(P)$ and $m - f(P)$. The first is non-negative in domain D, and the second non-positive. Hence (see property III),

$$\iint_D [M - f(P)]\,d\sigma \geqslant 0 \quad \text{and} \quad \iint_D [m - f(P)]\,d\sigma \leqslant 0,$$

i.e. $$\iint_D M\,d\sigma \geqslant \iint_D f(P)\,d\sigma \quad \text{and} \quad \iint_D m\,d\sigma \leqslant \iint_D f(P)\,d\sigma,$$

whence

$$m\iint\limits_D d\sigma \leqslant \iint\limits_D f(P)\,d\sigma \leqslant M\iint\limits_D d\sigma.$$

But these are the required inequalities, since integral $\iint\limits_D d\sigma$ is equal to the area S of D; in fact,

$$\iint\limits_D d\sigma = \lim \sum \Delta\sigma_i = \lim S = S.$$

We can thus find upper and lower bounds for a double integral without evaluating it, provided we know the maximum and minimum values of the integrand.

Property V may be seen geometrically by the obvious fact that the volume of a cylindrical solid is greater than the volume of the cylinder with the same base and height equal to the least z coordinate of the bounding surface, and less than the volume of the cylinder with the same base and height equal to the greatest z coordinate of the boundary surface.

REMARK 1. A more general theorem holds: *inequalities between functions imply inequalities in the same sense between their double integrals, i.e. if we have*

$$\varphi(P) \leqslant f(P) \leqslant \psi(P),$$

at any point P of domain D, we also have

$$\iint\limits_D \varphi(P)\,d\sigma \leqslant \iint\limits_D f(P)\,d\sigma \leqslant \iint\limits_D \psi(P)\,d\sigma;$$

in brief, *inequalities can be integrated.*

REMARK 2. The following inequality holds for the absolute value of a double integral:

$$\left|\iint\limits_D f(P)\,d\sigma\right| \leqslant \iint\limits_D \left|f(P)\right| d\sigma,$$

i.e. if $|f(P)| \leqslant M$ in domain D, we have

$$\left|\iint\limits_D f(P)\,d\sigma\right| \leqslant MS.$$

We leave the proofs of these two remarks to the reader.

170. Fundamental Properties of Double and Triple Integrals (continued). Additive Functions of a Domain. The Newton-Leibniz Formula.

We shall prove the following theorem.

THEOREM VI (on the mean value). **The double integral of a continuous function is equal to the product of the value of the function at**

some mean point of the domain of integration with the area of the domain.

Proof. By the theorem on the upper and lower bounds of an integral (property V):

$$mS \leqslant \iint_D f(P)\,d\sigma \leqslant MS$$

or

$$m \leqslant \frac{1}{S}\iint_D f(P)\,d\sigma \leqslant M.$$

At the same time, since (see Sec. 142) function $f(P)$ must take at some point $P_c(\xi, \eta)$ of domain D a value equal to a number intermediate to the minimum and maximum value of the function, we have

$$\frac{1}{S}\iint_D f(P)\,d\sigma = f(P_c). \qquad (*)$$

Hence

$$\iint_D f(P)\,d\sigma = f(P_c)\,S = f(\xi, \eta)\,S.$$

This is what we wished to prove.

The value $f(P_c)$ given by formula (*) is called the *mean value of* $f(P)$ *in domain* D.

Definition. **The *mean value* of a function of two independent variables in a domain is the ratio of the integral of the function over the domain to the area of the domain** (cf. Sec. 91).

The mean value for a double integral can be interpreted geometrically as follows: there exists a cylinder whose base is the same as base D of the given cylindrical body, whose height is equal to the z co-ordinate of surface $z = f(x, y)$ at some point of the base, and whose volume is equal to the volume of the cylindrical body. The z co-ordinate in question is the mean value of function $f(x, y)$ in domain D.

An important property of the ordinary integral is that its derivative with respect to the upper limit is equal to the integrand. We found the derivative of the integral (Sec. 92) by taking the limit of the ratio of the integral from a given point $P(x)$ over an interval of length Δl $(= \Delta x)$ to Δl itself as $\Delta l \to 0$. We can thus say briefly that:

Differentiation of an integral with respect to the interval over which it is taken yields the integrand.

The double integral has a similar property. Before turning to this property we must deal with the concept of additive function of a domain.

Definition. **A quantity u is called an *additive function of a variable domain* σ in a given domain D if a definite value of u corresponds to each domain σ belonging to D:**

$$u = F(\sigma);$$

and if, further, any domain σ is divided into non-overlapping domains $\sigma_1, \sigma_2, \ldots, \sigma_n$, so that* $\sigma = \sigma_1 + \sigma_2 + \cdots \sigma_n$, then

$$F(\sigma) = F(\sigma_1) + F(\sigma_2) + \cdots + F(\sigma_n).$$

Examples of *additive functions of a domain* include such frequently encountered quantities as length, area, volume, mass, work and so on. The temperature of a body (the same at all points) or the electrical tension in a conductor (the same at all points) provide examples of quantities that are not additive functions of the respective domains.

An infinitesimal calculus can be constructed for additive functions of a domain which is completely analogous to that for functions of one variable.

We shall define, in particular, the concepts of derivative and differential of an additive function $u = F(\sigma)$ of domain σ.

We take any neighbourhood $\Delta\sigma$ of point P in the domain in which the function is defined. We also use $\Delta\sigma$ to denote the area of this neighbourhood. We divide the value of the function $F(\Delta\sigma)$ corresponding to the neighbourhood by its area: $F(\Delta\sigma)/\Delta\sigma$, and pass to the limit on condition that domain $\Delta\sigma$ contracts to point $P(\Delta\sigma \to 0)$. If the limit exists, it is called the derivative of function $F(\sigma)$ with respect to domain σ at point P and is written as $F'(\sigma)$:

$$\lim_{\Delta\sigma \to 0} \frac{F(\Delta\sigma)}{\Delta\sigma} = F'(\sigma).$$

This derivative is a function of point P and not of domain σ.

Definition. **The principal part of the quantity $F(\Delta\sigma)$, i.e. of the quantity equivalent to $F(\Delta\sigma)$ as $\Delta\sigma \to 0$, which is linear in the area $\Delta\sigma$, is called the *differential* of function $F(\sigma)$ at point P and is written as $dF(\sigma)$.**

* The sum here implies taking the aggregate of all the pieces of plane belonging to all the "component" domains.

The reader will easily show that the following proposition holds as in the ordinary case:

The differential of $F(\sigma)$ is equal to the derivative of $F(\sigma)$ with respect to the domain multiplied by the differential of domain σ (area $\Delta\sigma = d\sigma$):

$$dF(\sigma) = F'(\sigma)\,d\sigma.$$

It follows from this that

$$F'(\sigma) = \frac{dF(\sigma)}{d\sigma}.$$

An additive function of a domain is said to be *differentiable* if it has a differential, or what amounts to the same thing, a derivative.

The double integral is an important example of an additive and differentiable function of a domain.

Let $z = f(P)$ be a function of two independent variables which is continuous in a plane domain D. We take the double integral of $f(P)$ over a domain σ lying wholly in D. Obviously, a definite value of the integral corresponds to every domain σ, i.e. the integral is a *function of the domain** of integration σ; we denote this function by $I(\sigma)$:

$$I(\sigma) = \iint_{\sigma} f(P)\,d\sigma.$$

Function $I(\sigma)$ is additive since it follows from property IV that, if domain σ is divided into (non-overlapping) sub-domains σ_1, σ_2, ..., σ_n, we have

$$I(\sigma) = I(\sigma_1) + I(\sigma_2) + \cdots + I(\sigma_n).$$

THEOREM VII (on the derivative of an integral with respect to the domain). **The derivative of a double integral with respect to the domain over which it is taken is equal to the integrand:**

$$\boldsymbol{I'(\sigma) = \frac{d}{d\sigma}\iint_{\sigma} f(P)\,d\sigma = f(P).} \qquad (*)$$

Proof. We take a neighbourhood $\Delta\sigma$ (with area $\Delta\sigma$) of a point P of domain D. We have:

$$I(\Delta\sigma) = \iint_{\Delta\sigma} f(P)\,d\sigma.$$

* The double integral with variable domain of integration corresponds to the ordinary integral with variable upper limit. The latter can also be regarded as an integral with a variable domain (interval) of integration.

By the mean value theorem:

$$\frac{I(\Delta\sigma)}{\Delta\sigma} = \frac{\iint\limits_{\Delta\sigma} f(P)\,d\sigma}{\Delta\sigma} = f(P_c),$$

where P_c is the "mean" point of domain $\Delta\sigma$. As $\Delta\sigma \to 0$, $P_c \to P$. Consequently, we have in view of the continuity of function $f(P)$:

$$\lim_{\Delta\sigma\to 0} \frac{I(\Delta\sigma)}{\Delta\sigma} = f(P).$$

This is what we needed to prove.

We have from expression (*) for the double integral:

$$dI(\sigma) = d\iint\limits_{\sigma} f(P)\,d\sigma = f(P)\,d\sigma,$$

i.e. *the element of the double integral* (the integrand element) *is the differential of the integral.* We can write conditionally:

$$I(D) = \iint\limits_{D} dI(\sigma).$$

Conversely, let an additive function u of domain σ, $u = F(\sigma)$, be previously assigned, having a continuous derivative $F'(\sigma) = f(P)$; as in the ordinary case, *the value of $F(\sigma)$ in domain D is now equal to the double integral over this domain of the differential of $F(\sigma)$:*

$$F(D) = \iint\limits_{D} dF(\sigma) = \iint\limits_{D} f(P)\,d\sigma. \qquad (**)$$

This proposition can be expressed as follows.

THE NEWTON–LEIBNIZ THEOREM. **A function of a domain possessing a continuous derivative is the double integral over the domain of its derivative.**

We shall not dwell on the proof.

Formula (**) is entirely analogous to the Newton–Leibniz formula for a single integral. We shall therefore refer to this as the *Newton–Leibniz formula for double integrals.*

The results of Secs. 169 and 170 can be carried over directly to triple integrals. The only changes required for the proofs are of terminology.

2. Iterated Integration

171. Evaluation of Double Integrals (Rectangular Domain). We divide the plane domain of integration D, referred to a system of Cartesian co-ordinates Oxy, into partial domains by means of two systems of *co-ordinate lines:* $x = \text{const.}$, $y = \text{const.}$ ($x = x_0, x_1, \ldots, x_n$; $y = y_0, y_1, \ldots, y_n$). These lines are straight lines parallel to Oy and Ox respectively, the partial domains being rectangles with sides parallel to the axes. Obviously, the area of a partial domain is $\Delta\sigma = \Delta x\, \Delta y$, and an elementary area $d\sigma$ is given by

$$d\sigma = dx\, dy,$$

i.e. *the differential of an area in rectangular Cartesian co-ordinates is equal to the product of the differentials of the independent variables.* We have*

$$I = \iint\limits_D f(P)\, d\sigma = \iint\limits_D f(x, y)\, dx\, dy = \lim_{\substack{m\to\infty \\ n\to\infty}} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} f(P_{ij})\, \Delta x_i\, \Delta y_j.$$

We shall base our evaluation of the double integral on the fact that (see Sec. 169) *every* double integral

$$I = \iint\limits_D f(P)\, dx\, dy$$

can be interpreted as the (algebraic) "volume" of a *cylindrical body* with base D bounded by the surface

$$z = f(P) = f(x, y).$$

We shall now evaluate this volume by the method indicated in Sec. 119.

We suppose first that the domain of integration is a *rectangle* $\overline{D}$ with sides parallel to the axes:

$$a \leqslant x \leqslant b,$$

$$c \leqslant y \leqslant d.$$

We consider the corresponding cylindrical body (there is no loss of generality in locating it in the first octant, as illustrated in

* The symbol $\sum_{i=0}^{n-1} \sum_{j=0}^{m-1}$ of "double summation" used here implies that we have to take first the sum of all the terms with $j = 0, 1, 2, \ldots, m-1$ and arbitrary i, then summate the sums obtained with $i = 0, 1, 2, \ldots, n-1$.

Fig. 31). We draw a plane parallel to Oyz and at some arbitrary distance $x = \text{const}$ ($a \leqslant x \leqslant b$) from it. This plane cuts the cylinder in the curvilinear trapezium $ABB'A'$ bounded by the plane curve L:

$$z = f(x, y), \quad x = \text{const}.$$

The area of trapezium $ABB'A'$ is given by the integral of the height of curve L over the base of the trapezium ($c \leqslant y \leqslant d$), i.e.

$$\int_c^d f(x, y)\,dy.$$

The integration is carried out with respect to y, the second argument x of the integrand being meantime reckoned constant.

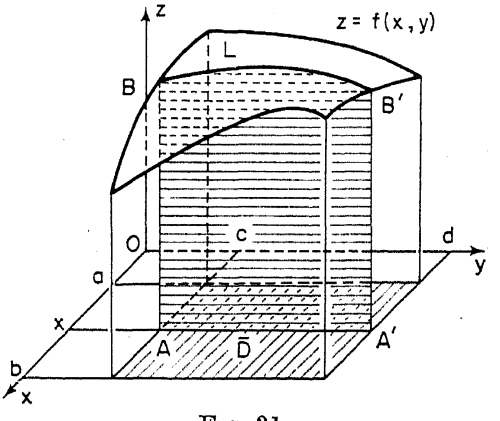

FIG. 31

The value of the integral just written depends on the value taken for x (i.e. on the distance of the cutting plane from Oyz); in other words, the integral is a function of x; we shall write this as $F(x)$:

$$F(x) = \int_c^d f(x, y)\,dy.$$

The volume of our solid is got by integration of the expression $F(x)$ for the area of the plane section over the interval of variation of x ($a \leqslant x \leqslant b$) (Sec. 119):

$$I = \int_a^b F(x)\,dx.$$

On equating this expression to the double integral, we get

$$\iint_{\bar{D}} f(x, y)\,dx\,dy = \int_a^b F(x)\,dx.$$

We find on substituting the expression for $F(x)$:

$$\iint_{\overline{D}} f(x, y)\, dx\, dy = \int_a^b \left(\int_c^d f(x, y)\, dy \right) dx,$$

or, as we usually write it,

$$\iint_{\overline{D}} f(x, y)\, dx\, dy = \int_a^b dx \int_c^d f(x, y)\, dy. \tag{A}$$

On the other hand, the "volume" can also be found from the areas of plane sections $ABB'A'$ (Fig. 32) parallel to Oxz (and not

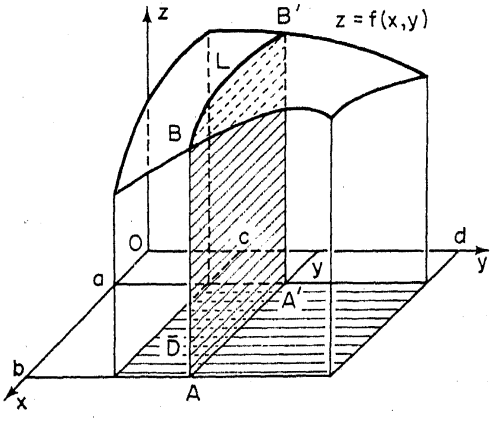

FIG. 32

Oyz). The area of a section by the plane $y = \text{const}$ is given by the integral

$$\Phi(y) = \int_a^b f(x, y)\, dx,$$

and the total "volume" by the integral of $\varphi(y)$ over the interval of variation of y ($c \leqslant y \leqslant d$). This leads to the formula

$$\iint_{\overline{D}} f(x, y)\, dx\, dy = \int_c^d \Phi(y)\, dy.$$

We can write in the same way as above:

$$\iint_{\overline{D}} f(x, y)\, dx\, dy = \int_c^d \left(\int_a^b f(x, y)\, dx \right) dy,$$

or

$$\iint_{\overline{D}} f(x, y)\, dx\, dy = \int_c^d dy \int_a^b f(x, y)\, dx. \tag{B}$$

On combining the two cases, we arrive at the following:

RULE FOR EVALUATION OF DOUBLE INTEGRALS. **To evaluate a double integral over a rectangle $\overline{D}$ we have to integrate the function with respect to one variable between the limits of its variation, then integrate the result with respect to the other variable between its limits of variation.**

This shows that a double integral can be evaluated with the aid of ordinary integrations in the case of a rectangular domain of integration.

Definition. **The right-hand sides of expressions (A) and (B) containing two successive operations of ordinary integration on function $f(x,y)$ are described as *iterated integrals of* $f(x, y)$ *in domain* $\overline{D}(a \leqslant x \leqslant b,\ c \leqslant y \leqslant d)$.**

In accordance with this terminology an ordinary integral is also described as single.

Equating (A) and (B) gives us

$$\int_c^d dy \int_a^b f(x, y)\,dx = \int_a^b dx \int_c^d f(x, y)\,dy,$$

i.e. **a (twice) iterated integral with constant limits of a continuous function does not depend on the order of integration.**

We note that, if the integrand is the product of two functions, each of which depends only on one variable:

$$f(x, y) = f_1(x)\, f_2(y),$$

the double integral over rectangle $\overline{D}$ is equal to the product of two ordinary integrals:

$$\iint_{\overline{D}} f(x, y)\,dx\,dy = \int_a^b f_1(x)\,dx \int_c^d f_2(y)\,dy.$$

In fact, since

$$\iint_{\overline{D}} f(x, y)\,dx\,dy = \int_c^d dy \int_a^b f_1(x)\, f_2(y)\,dx,$$

the required result follows on taking $f_2(y)$ outside the symbol of the inner integral then integral $\int_a^b f_1(x)\,dx$ (as constant) outside the symbol of the outer integral.

Example 1. Let us find the double integral of the function

$$z = 1 - \frac{x}{3} - \frac{y}{4}$$

over the rectangular domain $\overline{D}$ $(-1 \leqslant x \leqslant 1,\ -2 \leqslant y \leqslant 2)$:

$$I = \iint_{\overline{D}} \left(1 - \frac{x}{3} - \frac{y}{4}\right) dx\, dy.$$

Integral I expresses geometrically the volume of the quadrangular prism whose base is rectangle $\overline{D}$, intersected by the plane (Fig. 33)

$$\frac{x}{3} + \frac{y}{4} + z = 1.$$

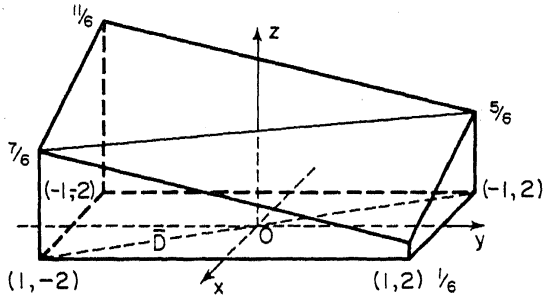

Fig. 33

We take the iterated integral, first with respect to x, then with respect to y:

$$I = \int_{-2}^{2} dy \int_{-1}^{1} \left(1 - \frac{x}{3} - \frac{y}{4}\right) dx = \int_{-2}^{2} \left(x - \frac{x^2}{6} - \frac{yx}{4}\right)\Big|_{-1}^{1} dy$$

$$= \int_{-2}^{2} \left(2 - \frac{y}{2}\right) dy = \left(2y - \frac{y^2}{4}\right)\Big|_{-2}^{2} = 8.$$

We get the same result by integrating first with respect to y then with respect to x:

$$I = \int_{-1}^{1} dx \int_{-2}^{2} \left(1 - \frac{x}{3} - \frac{y}{4}\right) dy = \int_{-1}^{1} \left(y - \frac{xy}{3} - \frac{y^2}{8}\right)\Big|_{-2}^{2} dx$$

$$= \int_{-1}^{1} \left(4 - \frac{4}{3}x\right) dx = \left(4x - \frac{2}{3}x^2\right)\Big|_{-1}^{1} = 8,$$

We can check our result by working out the volume by elementary methods. In fact (see Fig. 33):

$$I = \frac{4}{3}\left(\frac{1}{6} + \frac{5}{6} + \frac{7}{6}\right) + \frac{4}{3}\left(\frac{5}{6} + \frac{7}{6} + \frac{11}{6}\right) = \frac{4}{3} \cdot \frac{36}{6} = 8.$$

Example 2. Let us find the double integral of function $z = x^2 + y^2 - 2x - 2y + 4$ over the rectangular domain $\overline{D}$ $(0 \leqslant x \leqslant 2, 0 \leqslant y \leqslant 2)$:

$$I = \iint\limits_{\overline{D}} (x^2 + y^2 - 2x - 2y + 4)\, dx\, dy.$$

Geometrically, I gives the volume of the cylindrical body whose base is square $\overline{D}$ bounded by the paraboloid of revolution (Fig. 34):

$$z - 2 = (x - 1)^2 + (y - 1)^2.$$

The iterated integral, taken first with respect to x, then with respect to y, gives

$$\begin{aligned} I &= \int_0^2 dy \int_0^2 (x^2 + y^2 - 2x - 2y + 4)\, dx \\ &= \int_0^2 \left(\frac{x^2}{3} + y^2 x - x^2 - 2yx + 4x\right)\Bigg|_0^2 dy \\ &= \int_0^2 \left(\frac{20}{3} - 4y + 2y^2\right) dy \\ &= \left(\frac{20}{3} y - 2y^2 + \frac{2}{3} y^3\right)\Bigg|_0^2 = 10\,\frac{2}{3}. \end{aligned}$$

In view of the complete symmetry of x and y in the working it is unnecessary to verify that the result is the same if integration is first with respect to y, then with respect to x.

172. Evaluation of Double Integrals (Arbitrary Domain). Now let the domain of integration be any finite domain of the plane. We shall use similar arguments to show that the double integral is also given in this case by an iterated integral.

We first let the domain of integration satisfy the following condition: *any straight line parallel to Ox or Oy cuts the boundary of the domain in not more than two points.* We again use $\overline{D}$ for denoting a domain of this type. The general case reduces to this, as we shall see.

We enclose $\overline{D}$ in some rectangle

$$a \leqslant x \leqslant b, \quad c \leqslant y \leqslant d$$

(Fig. 35), the sides of which touch the boundary at points A, B, C, D. Interval $[a, b]$ is the orthogonal projection of $\overline{D}$ on Ox, and interval $[c, d]$ the orthogonal projection of $\overline{D}$ on Oy. It should be

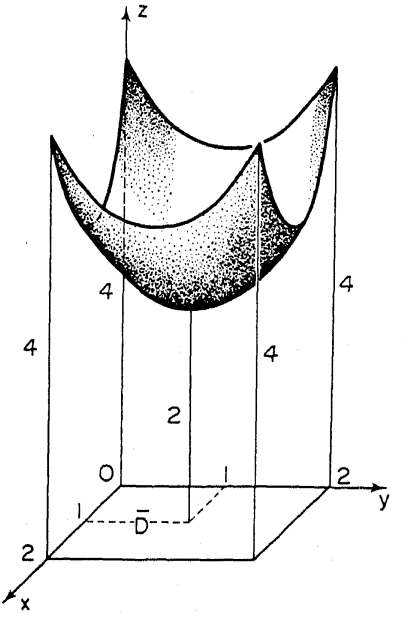

FIG. 34

noted that the boundary of $\overline{D}$ can also include pieces of straight lines (including those parallel to the axes).

Points A and C divide the boundary into two parts: ABC and ADC, each of which is cut by any line parallel to Oy in not more than *one* point. Their equations can therefore be written explicitly with regard to y:

$$y = \varphi_1(x) \quad (ABC),$$

$$y = \varphi_2(x) \quad (ADC),$$

where φ_1 and φ_2 are *single-valued* functions of x in interval $[a, b]$.

Similarly, points B and D divide the boundary into curves BAD and BCD, the equations of which can be written as

$$x = \psi_1(y) \quad (BAD),$$

$$x = \psi_2(y) \quad (BCD),$$

where ψ_1 and ψ_2 are *single-valued* functions of y in the interval $[c, d]$.

We cut the cylindrical body in question by an arbitrary plane parallel to Oyz, i.e. $x = \text{const.}$, $a \leqslant x \leqslant b$ (Fig. 36). This gives us the curvilinear trapezium $MNPR$, the area of which is given by the ordinary integral of function $f(x, y)$, regarded as a function of

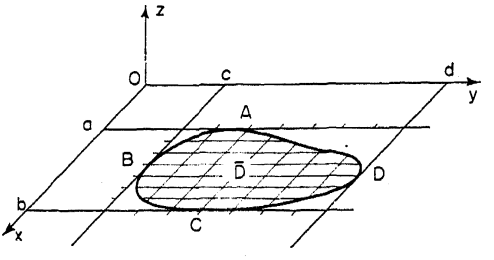

FIG. 35

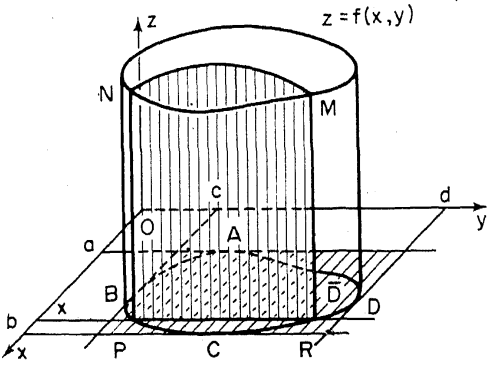

FIG. 36

the single variable y, varying from the ordinate of point P to the ordinate of R. Point P is where the line $x = \text{const.}$ "enters" domain $\overline{D}$ (in the plane Oxy), whilst the line "leaves" the domain at R. Since the equation of curve ABC is $y = \varphi_1(x)$, whilst curve ACD is $y = \varphi_2(x)$, these ordinates are $\varphi_1(x)$ and $\varphi_2(x)$ respectively with x equal to the constant in question. Consequently the integral

$$\int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\, dy$$

gives the area of plane section $MNPR$ in terms of the distance x of the cutting plane from the parallel plane Oyz. As in the case of a rectangular domain $\overline{D}$, the volume of the whole solid is equal to

the integral of this expression with respect to x over the interval of variation of x ($a \leqslant x \leqslant b$). Thus

$$\iint\limits_{\bar{D}} f(x, y)\, dx\, dy = \int\limits_{a}^{b} dx \int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\, dy. \qquad (*)$$

The inner integral (with respect to y) differs from that in the case of a rectangular domain (Sec. 171) in that the limits are *variable* and not constant, though their meaning remains the same; they indicate the bounds of variation of the variable of integration (y) for a constant value of the second argument (x). The limits of the outer integral (with respect to x) are constant; they indicate the bounds within which the second argument (x) can vary, x having been regarded as constant in the first integration.

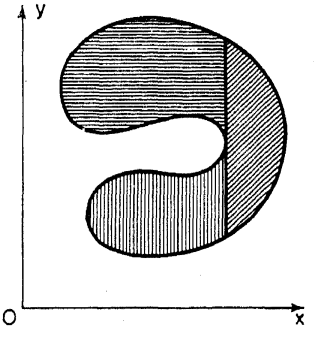

FIG. 37

By exchanging the roles of x and y, we arrive at a second expression for the double integral:

$$\iint\limits_{\bar{D}} f(x, y)\, dx\, dy = \int\limits_{c}^{d} dy \int\limits_{\psi_1(y)}^{\psi_2(y)} f(x, y)\, dx. \qquad (**)$$

The integration here is carried out first with respect to x, then with respect to y, whilst the variable limits of the inner integration indicate the bounds of variation of the variable of integration (x) for a constant value of the second argument (y). The limits of the outer integral (with respect to y) are constant; they indicate the bounds within which the second argument (y) can vary.

RULE FOR EVALUATION OF DOUBLE INTEGRALS. **To evaluate a double integral over a domain $\bar{D}$ we have to integrate the function with respect to one variable between the limits of its variation for a constant but arbitrary value of the second variable, then integrate the result with respect to the second variable, between the limits of its maximum variation in the domain of integration.**

The right-hand sides of expressions (*) and (**) are also called *iterated integrals of function $f(x, y)$ in domain $\bar{D}$*. Their strict definitions do not differ from those of Sec. 171 for an iterated integral.

We suppose finally that the domain of integration is any finite domain in the plane. As shown by Fig. 37, such a domain can be

divided into a number of domains of the form of $\overline{D}$. The double integral over the whole domain can then be written as the sum of the double integrals over the constituent domains, in accordance with property IV (Sec. 169). Each of these integrals reduces to an iterated integral, so that a double integral may be evaluated in the general case by means of a series of ordinary integrations.

The concept of *double* integral is nevertheless of great importance since it enables us to use *one* double integral for *any* plane domain instead of a *sum* of iterated integrals. The essential point is that the double integral has all the basic properties of the ordinary integral.

REMARK 1. *Area of a figure.* The double integral of unity over a domain D:

$$\iint\limits_D d\sigma = \iint\limits_D dx\,dy,$$

is numerically equal to the **area of domain** D. The area of the domain (figure) can thus be expressed by means of one double integral.

If D is the domain $\overline{D}$, its area S is given by the iterated integral

$$S = \int_a^b dx \int_{\varphi_1(x)}^{\varphi_2(x)} dy, \quad \text{or} \quad S = \int_c^d dy \int_{\psi_1(y)}^{\psi_2(y)} dx,$$

where the notation is as above. Hence

$$S = \int_a^b [\varphi_2(x) - \varphi_1(x)]\,dx, \quad \text{or} \quad S = \int_c^d [\psi_2(y) - \psi_1(y)]\,dy.$$

These expressions for the area of domain D also follow directly from the geometrical meaning of the single integral.

REMARK 2. *Changing the order of integration.* The fact that the same double integral over domain $\overline{D}$ gives two different iterated integrals leads to a rule for transforming iterated integrals. We have from equations (*) and (**):

$$\int_a^b dx \int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\,dy = \int_c^d dy \int_{\psi_1(y)}^{\psi_2(y)} f(x, y)\,dx.$$

This case differs from that of a rectangular domain $\overline{D}$ in that changing the order of integration implies changing the limits of integration. Hence a special formula for transforming iterated integrals is obtained for each concrete domain $\overline{D}$. Some of these formulae prove useful in various operations on integrals.

We shall mention one such formula, which refers to integration over a right-angled isosceles triangle. We consider the double integral of function $f(x, y)$ over the domain given by $a \leqslant x \leqslant b$, $a \leqslant y \leqslant x$ (Fig. 38). It may easily be found by integrating first with respect to y then with respect to x and equating the result to that got by integrating first with respect to x then with respect to y that

$$\int_a^b dx \int_a^x f(x, y)\, dy = \int_a^b dy \int_y^b f(x, y)\, dx.$$

This transformation formula for an iterated integral is known as *Dirichlet's formula*.

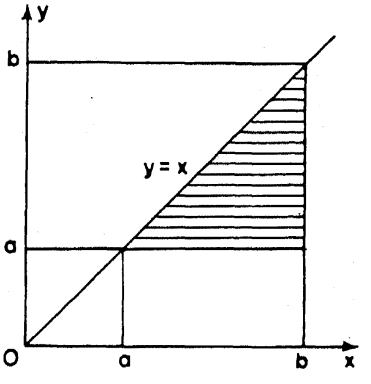

FIG. 38

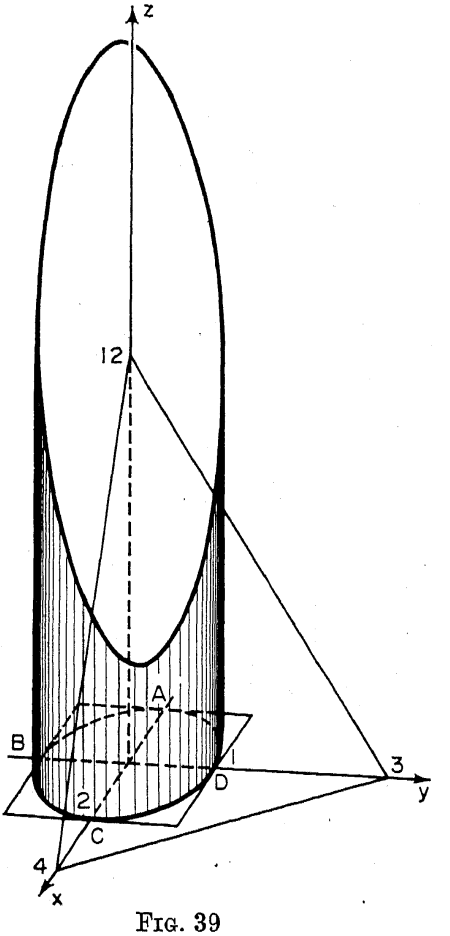

FIG. 39

Example 1. Let us evaluate the double integral of $z = 12 - 3x - 4y$ over domain $\overline{D}$ given by $x^2 + 4y^2 \leqslant 4$:

$$I = \iint_{\overline{D}} (12 - 3x - 4y)\, dx\, dy.$$

Geometrically, I is the volume of a cylinder whose base is the interior of the ellipse $x^2/4 + y^2 = 1$ cut by the plane

$$\frac{x}{4} + \frac{y}{3} + \frac{z}{12} = 1.$$

The truncated cylinder is illustrated in Fig. 39.

We first integrate with respect to x, then with respect to y. Since the equations of curves BAD and BCD are

$$x = -2\sqrt{1-y^2},$$

$$x = 2\sqrt{1-y^2},$$

x varies from $-2\sqrt{1-y^2}$ to $2\sqrt{1-y^2}$ when y is constant. As regards y, it can vary from -1 to 1. Thus

$$I = \int_{-1}^{1} dy \int_{-2\sqrt{1-y^2}}^{2\sqrt{1-y^2}} (12 - 3x - 4y)\, dx.$$

We have by the familiar property of integrals (see Sec. 107):

$$I = 8\int_{-1}^{1} dy \int_{0}^{2\sqrt{1-y^2}} (3-y)\, dx = 16\int_{-1}^{1} (3-y)\sqrt{1-y^2}\, dy;$$

we obtain further, on using the same property:

$$I = 96\int_{0}^{1} \sqrt{1-y^2}\, dy = 96 \cdot \frac{\pi}{4} = 24\pi.$$

The answer may be checked by integrating first with respect to y, then with respect to x. The equations of curves ABC and ADC are

$$y = -\sqrt{1-\frac{x^2}{4}}$$

and

$$y = \sqrt{1-\frac{x^2}{4}},$$

whilst x varies from -2 to 2. Therefore

$$I = \int_{-2}^{2} dx \int_{-\sqrt{1-\frac{x^2}{4}}}^{\sqrt{1-\frac{x^2}{4}}} (12 - 3x - 4y)\, dy.$$

We now have:

$$I = 2\int_{-2}^{2} (12-3x)\, y \Big|_{0}^{\sqrt{1-\frac{x^2}{4}}} dx = 48\int_{0}^{2} \sqrt{1-\frac{x^2}{4}}\, dx$$

$$= 96\int_{0}^{1} \sqrt{1-t^2}\, dt = 96 \cdot \frac{\pi}{4} = 24\pi.$$

Example 2. Let us find the volume V of the body bounded by the surface

$$z = 1 - 4x^2 - y^2$$

and the plane Oxy. The body is a segment of an elliptic paraboloid situated above the Oxy plane (Fig. 40). The paraboloid is cut by Oxy in the ellipse

$$4x^2 + y^2 = 1.$$

The problem thus amounts to finding the volume of the cylindrical body having the interior of the ellipse as its base and bounded by

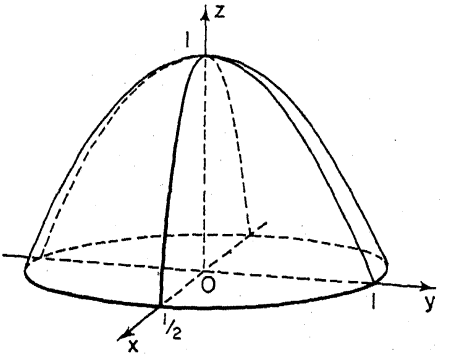

FIG. 40

the paraboloid $z = 1 - 4x^2 - y^2$. (This cylindrical body has no lateral cylindrical surface, just as the curvilinear trapezium bounded say by a sine wave in the interval $(0, \pi)$ has no sides.)

In view of the symmetry of the body with respect to the Oxz and Oyz planes, the volume contained in the first octant will be a quarter of the total; this is equal to the double integral over the domain given by

$$4x^2 + y^2 \leqslant 1, \quad x \geqslant 0, \quad y \geqslant 0.$$

Integration with respect to y, then with respect to x, gives

$$\frac{V}{4} = \int_0^{\frac{1}{2}} dx \int_0^{\sqrt{1-4x^2}} (1 - 4x^2 - y^2)\, dy = \frac{2}{3} \int_0^{\frac{1}{2}} (1 - 4x^2)^{\frac{3}{2}}\, dx.$$

We obtain by substituting $2x = \sin t$:

$$\frac{V}{4} = \frac{2}{3} \cdot \frac{1}{2} \int_0^{\frac{\pi}{2}} \cos^4 t\, dt = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{16} \pi$$

(see Sec. 106), whence

$$V = \frac{\pi}{4}.$$

Integrating in the reverse order leads to rather shorter working, as the reader may verify for himself. The fact that we can choose the order of the iterated integration is in fact generally used so as to obtain the simplest possible working.

173. Evaluation of Triple Integrals. Triple integrals can also be evaluated by means of a series of single integrations. We shall confine ourselves to describing the rule.

Suppose we are given the triple integral of function $f(P)$ over some finite domain Ω of space:

$$I = \iiint_{\Omega} f(P)\, dv,$$

Ω being referred to a system of Cartesian co-ordinates $Oxyz$.

We sub-divide Ω by planes parallel to the co-ordinate planes. The sub-domains will be parallelepipeds with faces parallel to the Oxy, Oxz and Oyz planes, and an elementary volume in Ω will be equal to the product of the differentials of the variables of integration: $dv = dx\, dy\, dz$. We write in accordance with this:

$$I = \iiint_{\Omega} f(x, y, z)\, dx\, dy\, dz.$$

Suppose that any straight line parallel to one of the axes cuts domain Ω *in not more than two points*. We shall use $\overline{\Omega}$ to denote such a domain. If this hypothesis is not true, we divide Ω so that each part of it is a domain $\overline{\Omega}$, and write the given integral as the sum of the integrals over the constituent domains.

We circumscribe a cylindrical surface perpendicular to Oxy about the domain (body) $\overline{\Omega}$ (Fig. 41). It touches $\overline{\Omega}$ along a curve L which divides the surface bounding the domain into two parts: an upper and a lower. Let the equation of the lower part be $z = \chi_1(x, y)$, and of the upper part $z = \chi_2(x, y)$.

The cylindrical surface cuts out a plane domain $\overline{D}$ from the Oxy plane, $\overline{D}$ being the orthogonal projection of spatial domain $\overline{\Omega}$ on the Oxy plane; curve L now projects into the boundary of $\overline{D}$. Functions $\chi_1(x, y)$ and $\chi_2(x, y)$ are single-valued in $\overline{D}$.

We shall integrate first over the Oz direction. This is done by integrating $f(x, y, z)$ over the straight segment contained in $\bar{\Omega}$ which is parallel to Oz and passes through some point $P(x, y)$ of domain $\bar{D}$ (segment $\alpha\beta$ in Fig. 41). The variable of integration z will vary, for given x and y, from $\chi_1(x, y)$ to $\chi_2(x, y)$; $\chi_1(x, y)$ is the z co-ordinate of point α (the point at which the straight line "enters" $\bar{\Omega}$), and $\chi_2(x, y)$ the z co-ordinate of β (the point at which

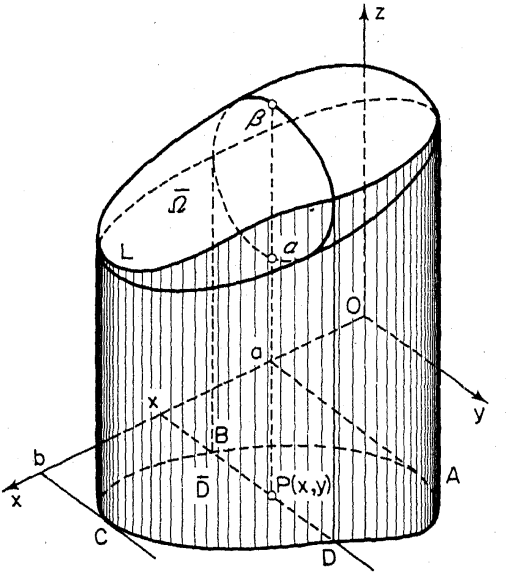

FIG. 41

the straight line leaves $\bar{\Omega}$). The result of integration is a magnitude depending on the point $P(x, y)$; let it be denoted by $F(x, y)$:

$$F(x, y) = \int_{\chi_1(x,y)}^{\chi_2(x,y)} f(x, y, z)\,dz.$$

We regard x and y as constants during the present integration.

We obtain the required triple integral by integrating $F(x, y)$ on the assumption that point $P(x, y)$ varies over domain $\bar{D}$, i.e. by finding the double integral

$$\iint_{\bar{D}} F(x, y)\,dx\,dy.$$

The triple integral I can therefore be written as

$$I = \iint_{\bar{D}} dx\,dy \int_{\chi_1(x,y)}^{\chi_2(x,y)} f(x, y, z)\,dz.$$

On further expressing the double integral over $\overline{D}$ as an iterated integral, first with respect to y, then with respect to x, we get

$$I = \int_a^b dx \int_{\varphi_1(x)}^{\varphi_2(x)} dy \int_{\chi_1(x,y)}^{\chi_2(x,y)} f(x, y, z)\, dz, \qquad (*)$$

where $\varphi_1(x)$ and $\varphi_2(x)$ are the ordinates of the points of "entry" and "departure" of the straight line $x = \text{const}$ into and from $\overline{D}$ (in the Oxy plane) and a, b are the abscissae of the ends of the interval of Ox representing the projection of $\overline{D}$.

The triple integral over domain $\overline{\Omega}$ is thus seen to be evaluated by means of three single integrations.

Definition. **The right-hand side of equation (*), consisting of three successive single integrations of function $f(x, y, z)$, is called *the (thrice) iterated integral of $f(x, y, z)$ over domain $\overline{\Omega}$.***

A different order can be chosen for carrying out the integrations. For instance, we can integrate first over the Oy direction, then over a domain in the Oxz plane formed by projection of the domain of integration $\overline{\Omega}$ on to Oxz. This leads to a change of order in the iterated integral, and in general to a change in the limits of integration for each variable.

RULE FOR EVALUATING TRIPLE INTEGRALS. **To find a triple integral over a domain Ω, we must:**

(1) integrate the function with respect to one variable between its limits of variation when the other two are arbitrarily held constant;

(2) integrate the result of the first integration with respect to a second variable between its limits of variation when the third is arbitrarily held constant; the domain of integration is the projection of the given domain Ω on the plane of the second and third variables;

(3) integrate the last result with respect to the third variable between the limits of its maximum variation in the plane domain concerned.

If the domain of integration $\overline{\Omega}$ is the interior of a parallelepiped with faces parallel to the co-ordinate planes (Fig. 42), the limits in all three integrals will be constants and will remain the same when the order of integration is altered:

$$I = \iiint_{\Omega} f(P)\, dx\, dy\, dz = \int_a^b dx \int_c^d dy \int_k^l f(P)\, dz$$

$$= \int_c^d dy \int_a^b dx \int_k^l f(P)\, dz = \cdots$$

REMARK. The triple integral of unity over a domain Ω of space:

$$I = \iiint_{\Omega} dV = \iiint_{\Omega} dx\, dy\, dz,$$

is numerically equal to the *volume V of domain Ω*. For,

$$I = \lim \sum_{i=1}^{n} \Delta v_i = \lim V = V.$$

The volume of *any* spatial domain (body) can therefore be given by one triple integral.

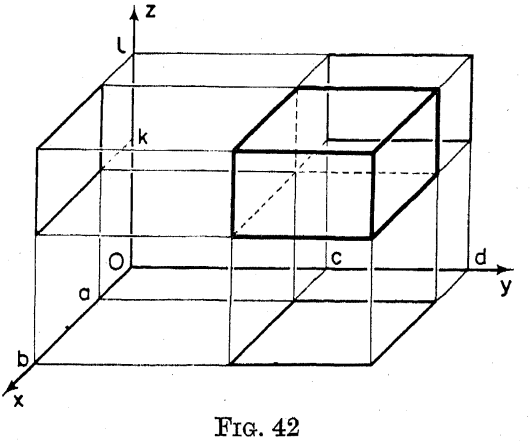

FIG. 42

Example. Let us find the volume V of the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

We have:

$$V = \iiint_{\Omega} dx\, dy\, dz,$$

where Ω is the spatial domain bounded by the ellipsoid.

We establish the order of integration for the iterated integral: first with respect to z, then with respect to y, then with respect to x. Obviously, with constant (but arbitrary) x and y, z varies from $-c\sqrt{1 - x_2/a^2 - y^2/b^2}$ to $c\sqrt{1 - x^2/a^2 - y^2/b^2}$; since the projection of Ω on to the Oxy plane is the interior of the ellipse $x^2/a^2 + y^2/b^2 = 1$, given constant x, y varies from $-b\sqrt{1 - x^2/a^2}$ to $b\sqrt{1 - x^2/a^2}$; finally, x can vary from $-a$ to a. Consequently,

$$V = \int_{-a}^{a} dx \int_{-b\sqrt{1-\frac{x^2}{a^2}}}^{b\sqrt{1-\frac{x^2}{a^2}}} dy \int_{-c\sqrt{1-\frac{x^2}{a^2}-\frac{y^2}{b^2}}}^{c\sqrt{1-\frac{x^2}{a^2}-\frac{y^2}{b^2}}} dz.$$

Hence

$$V = \int_{-a}^{a} dx \int_{-b\sqrt{1-\frac{x^2}{a^2}}}^{b\sqrt{1-\frac{x^2}{a^2}}} 2c \sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}}\, dy.$$

We find on putting $y = b\sqrt{1 - x^2/a^2}\sin t$:

$$V = \int_{-a}^{a} 4bc\left(1 - \frac{x^2}{a^2}\right) dx \int_{0}^{\frac{\pi}{2}} \cos^2 t\, dt = \pi bc \int_{-a}^{a}\left(1 - \frac{x^2}{a^2}\right) dx$$

$$= \frac{4}{3}\pi abc.$$

3. Integrals in Polar, Cylindrical and Spherical Co-ordinates

174. The Double Integral in Polar Co-ordinates.

I. We have so far used a system of Cartesian co-ordinates in a plane for evaluating a double integral over a domain D. We now refer configurations in the plane to a system of polar co-ordinates (ϱ, φ)

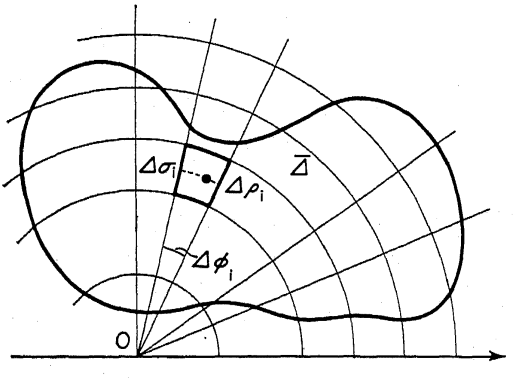

FIG. 43

and divide the domain of integration D into sub-domains σ_i by two systems of *co-ordinate lines* $\varrho = \text{const.}$, $\varphi = \text{const.}$: $\varrho = \varrho_1, \varrho_2, \ldots, \varrho_n$; $\varphi = \varphi_1, \varphi_2, \ldots, \varphi_m$, the former being concentric circles with centres at the pole and the latter radius vectors from the pole (Fig. 43). Sub-domains σ_i are curvilinear quadrilaterals, bounded

by two concentric circles and their radii. We have for the area $\Delta\sigma_i$ of domain $\Delta\sigma_i$:

$$\Delta\sigma_i = \frac{1}{2}(\varrho_i + \Delta\varrho_i)^2\Delta\varphi_i - \frac{1}{2}\varrho_i^2\,\Delta\varphi_i = \left(\varrho_i + \frac{\Delta\varrho_i}{2}\right)\Delta\varrho_i\Delta\varphi_i,$$

or

$$\Delta\sigma_i = \varrho_i'\Delta\varrho_i\Delta\varphi_i,$$

where

$$\varrho_i' = \frac{\varrho_i + (\varrho_i + \Delta\varrho_i)}{2}$$

is the mean radius between ϱ_i and $\varrho_i + \Delta\varrho_i$.

Let $F(P)$ be a continuous function in domain D. On choosing as points P_i in the integral sum

$$I_n = \sum_{i=1}^{n} F(P)_i\Delta\sigma_i$$

arbitrary points of domain σ_i lying on the mean circles $\varrho = \varrho_i'$, we get

$$I_n = \sum_{i=1}^{n} F(\varrho_i', \varphi_i')\varrho_i'\Delta\varrho_i\Delta\varphi_i.$$

But the right-hand side is an integral sum for function $F(\varrho, \varphi)$ with respect to variables ϱ, φ (Sec. 167). Passage to the limit gives us

$$I = \iint_D F(P)d\sigma = \iint_D F(\varrho, \varphi)\,\varrho\,d\varrho\,d\varphi.$$

*The elementary (differential) area in polar co-ordinates is**:

$$d\sigma = \varrho\,d\varrho\,d\varphi.$$

Given the double integral

$$\iint_D f(x, y)\,dx\,dy,$$

if we replace the Cartesian by the polar co-ordinates of a point, $x = \varrho\cos\varphi$, $y = \varrho\sin\varphi$, we arrive at the equation

$$\iint_D f(x, y)\,dx\,dy = \iint_D F(\varrho, \varphi)\,\varrho\,d\varrho\,d\varphi,$$

where $F(\varrho, \varphi) = f(\varrho\cos\varphi, \varrho\sin\varphi)$.

* The expression for $d\sigma$ is found at once by neglecting higher order infinitesimals and taking σ_i to be a rectangle with sides $\varrho_i\delta_i\varphi$ and $\delta\varrho_i$.

This is the *expression for transforming a double integral from Cartesian to polar co-ordinates.*

RULE FOR TRANSFORMING A DOUBLE INTEGRAL TO POLAR CO-ORDINATES. **To transform a double integral in Cartesian co-ordinates to a double integral in polar co-ordinates, we have to replace x and y in the integrand by $\varrho \cos \varphi$ and $\varrho \sin \varphi$ respectively, multiply the result by ϱ and take the product of differentials $d\varrho\, d\varphi$ instead of $dx\, dy$.**

II. The question naturally arises of evaluating a double integral in polar co-ordinates with the aid of single integrations over ϱ and φ.

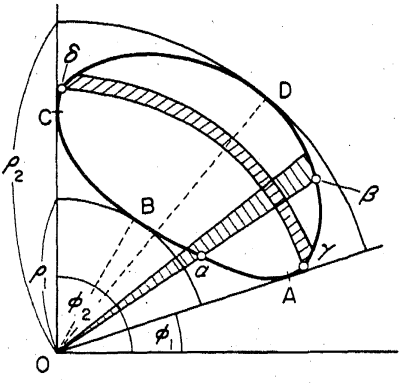

FIG. 44

1. If the domain of integration does not contain the pole (origin) and the co-ordinate lines $\varrho = \text{const.}$, $\varphi = \text{const.}$ meet its boundary in not more than two points (we shall refer to such a domain as a $\overline{\Delta}$ domain), the double integral in polar co-ordinates reduces to two different iterated integrals precisely as in the case of Cartesian co-ordinates.

We enclose the $\overline{\Delta}$ domain inside a curvilinear quadrilateral formed by the curves $\varrho = \varrho_1$, $\varrho = \varrho_2$, $\varphi = \varphi_1$, $\varphi = \varphi_2$ (Fig. 44). Domain $\overline{\Delta}$ touches the quadrilateral in points A, B, C, D. We shall first "collect" the elements of the integral along the radius vector $\varphi = \text{const}$, then "sum" the results obtained over the polar angle; in other words, we first integrate the function $F(\varrho, \varphi)\varrho$ with respect to ϱ between the limits of its variation for constant but arbitrary φ, then integrate the result with respect to φ between the limits of its maximum variation from φ_1 to φ_2:

$$I = \int_{\varphi_1}^{\varphi_2} d\varphi \int_{\nu_1(\varphi)}^{\nu_2(\varphi)} F(\varrho, \varphi)\, \varrho\, d\varrho, \qquad (*)$$

where $\varrho = \nu_1(\varphi)$ and $\varrho = \nu_2(\varphi)$ are the polar equations of curves ABC and ADC respectively ($\nu_1(\varphi)$ and $\nu_2(\varphi)$ are *single-valued* functions of φ in the interval $[\varphi_1, \varphi_2]$). Here $\nu_1(\varphi)$ is the radius vector of α, the point of "entry" of the straight line $\varphi = \text{const.}$ into domain $\overline{\Delta}$, and $\nu_2(\varphi)$ is the radius vector of β, the point of "departure" of the line from $\overline{\Delta}$.

Summation of the elements in the reverse order – first over the circle $\varrho = \text{const.}$ and then over the radius vector – is used more rarely; it leads to the formula

$$I = \int_{\varrho_1}^{\varrho_2} \varrho \, d\varrho \int_{\mu_1(\varrho)}^{\mu_2(\varrho)} F(\varrho, \varphi) \, d\varphi ,$$

where $\varphi = \mu_1(\varrho)$ and $\varphi = \mu_2(\varrho)$ are the polar equations of curves BAD and BCD respectively ($\mu_1(\varrho)$ and $\mu_2(\varrho)$ are *single-valued* functions of ϱ in the interval $[\varrho_1, \varrho_2]$). Here $\mu_1(\varrho)$ is the polar angle of γ, the point of "entry" of circle $\varrho = \text{const}$ into $\overline{\Delta}$; $\mu_2(\varrho)$ is the polar angle of δ, the point of "departure" of the circle from $\overline{\Delta}$.

In the particular case when the domain of integration is the interior of the quadrilateral $\varrho_1 \leqslant \varrho \leqslant \varrho_2$, $\varphi_1 \leqslant \varphi \leqslant \varphi_2$, the limits are constant and do not change With the order of integration:

$$I = \int_{\varphi_1}^{\varphi_2} d\varphi \int_{\varrho_1}^{\varrho_2} F(\varrho, \varphi) \, \varrho \, d\varrho = \int_{\varrho_1}^{\varrho_2} \varrho \, d\varrho \int_{\varphi_1}^{\varphi_2} F(\varrho, \varphi) \, d\varphi .$$

2. If the domain of integration contains the pole and any radius vector cuts its boundary in a single point (it is said to be star-shaped in this case with respect to the pole), we find by integrating first with respect to ϱ then with respect to φ:

$$I = \int_0^{2\pi} d\varphi \int_0^{\nu(\varphi)} F(\varrho, \varphi) \, \varrho \, d\varrho ,$$

where $\varrho = \nu(\varphi)$ is the polar equation of the boundary.

In particular, when $\varrho = \nu(\varphi) = R$, i.e. the domain of integration is a circle with centre at the pole, we have

$$I = \int_0^{2\pi} d\varphi \int_0^{R} F(\varrho, \varphi) \, \varrho \, d\varrho .$$

If the domain of integration of a double integral that is to be reduced to single integrals with respect to ϱ and φ does not satisfy the above conditions it has to be sub-divided so that each part satisfies the conditions.

It may be observed finally that, when $F(\varrho, \varphi) = 1$, the double integral in polar co-ordinates is numerically equal to the area S of the domain of integration Δ:

$$S = \iint_{\Delta} \varrho \, d\varrho \, d\varphi.$$

We obtain from (*) for the area of Δ:

$$S = \int_{\varphi_1}^{\varphi_2} d\varphi \int_{\nu_1(\varphi)}^{\nu_2(\varphi)} \varrho \, d\varrho = \frac{1}{2} \int_{\varphi_1}^{\varphi_2} [\nu_2^2(\varphi) - \nu_1^2(\varphi)] \, d\varphi;$$

in particular, we have for the area of the curvilinear sector ($\nu_1(P) = 0$, $\nu_2(\varphi) = \nu(\varphi)$):

$$S = \frac{1}{2} \int_{\varphi_1}^{\varphi_2} \nu^2(\varphi) \, d\varphi,$$

i.e. the formula obtained in Sec. 117.

Example. Let us find the volume V of the part common to a sphere of radius a and a circular cylinder of radius $\frac{1}{2}a$ passing

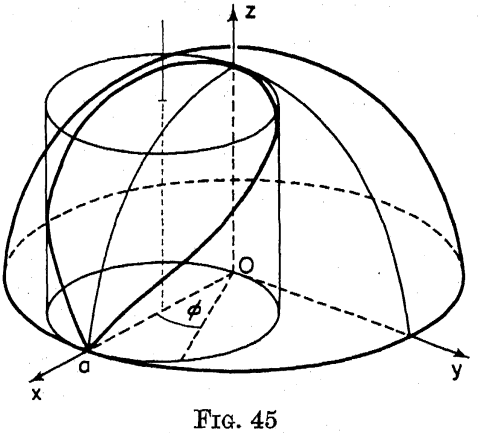

FIG. 45

through the centre of the sphere (known as Viviani's problem (1622–1703)). We arrange the co-ordinate system as shown in Fig. 45. In view of the symmetry of the required volume with respect to the Oxy and Oxz planes a quarter of the volume will be located in the first octant.

We have

$$\frac{1}{4} V = \iint_{D} \sqrt{a^2 - x^2 - y^2} \, dx \, dy,$$

where D denotes the interior of half the base of the cylinder.

The double integral may be conveniently transformed to polar co-ordinates. We have by the transformation rule:

$$\frac{1}{4}V = \iint_D \sqrt{a^2 - \varrho^2}\,\varrho\,d\varrho\,d\varphi.$$

Since the polar equation of the semi-circle bounding D is $\varrho = a\cos\varphi$, we find by integrating first with rspect to ϱ, then with respect to φ:

$$\frac{1}{4}V = \int_0^{\frac{\pi}{2}} d\varphi \int_0^{a\cos\varphi} \sqrt{a^2 - \varrho^2}\,\varrho\,d\varrho.$$

Evaluation of the inner integral gives us

$$\frac{1}{4}V = \frac{a^2}{3}\int_0^{\frac{\pi}{2}} (1 - \sin^3\varphi)\,d\varphi = \frac{a^3}{3}\left(\frac{\pi}{2} - \frac{2}{3}\right).$$

Hence

$$V = \frac{4}{3}a^3\left(\frac{\pi}{2} - \frac{2}{3}\right).$$

An interesting point is that the volume of the remainder of the hemisphere (after removing V) is expressed rationally in terms of the radius a of the sphere. The volume is, in fact:

$$\frac{2}{3}\pi a^3 - \frac{4}{3}a^3\left(\frac{\pi}{2} - \frac{2}{3}\right) = \frac{8}{9}a^3.$$

175. Triple Integrals in Cylindrical and Spherical Co-ordinates. Let us consider the triple integral of function $F(P)$ continuous in domain Ω:

$$I = \iiint_\Omega F(P)\,dv.$$

We shall find the expressions for I in cylindrical and spherical co-ordinates.

I. CYLINDRICAL CO-ORDINATES. We refer domain Ω *to a system of cylindrical co-ordinates* (ϱ, φ, z) in which the position of a point P in space is given by the polar co-ordinates (ϱ, φ) of its projection on the Oxy plane and by its z co-ordinate.

We divide the domain Ω into sub-domains v_i by means of the three systems of *co-ordinate surfaces:* $\varrho = \text{const.}$, $\varphi = \text{const.}$, $z = \text{const.}$, these being respectively circular cylindrical surfaces with axis Oz, half-planes through Oz, and planes parallel to Oxy. Sub-domains v_i consist of right cylinders MN (Fig. 46). Since the

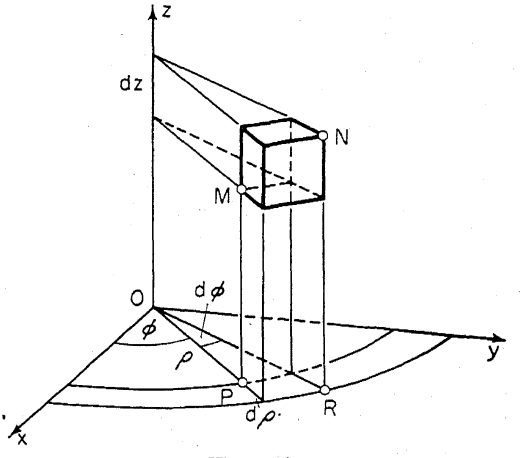

FIG. 46

volume of cylinder MN is equal to the base area multiplied by the height, we obtain for an elementary volume:

$$dv = \varrho \, d\varrho \, d\varphi \, dz,$$

and

$$I = \iiint_{\Omega} F(P) \, dv = \iiint_{\Omega} F(\varrho, \varphi, z) \, \varrho \, d\varrho \, d\varphi \, dz.$$

A triple integral given in Cartesian co-ordinates is readily transformed to cylindrical co-ordinates by bearing in mind that the Cartesian and cylindrical co-ordinates of a point P in space are connected for the system shown in Fig. 46 by the equations $x = \varrho \cos\varphi$, $y = \varrho \sin\varphi$, $z = z$; in fact,

$$I = \iiint_{\Omega} f(x, y, z) \, dx \, dy \, dz = \iiint_{\Omega} F(\varrho, \varphi, z) \, \varrho \, d\varrho \, d\varphi \, dz,$$

where

$$F(\varrho, \varphi, z) = f(\varrho \cos\varphi, \varrho \sin\varphi, z).$$

When $F(\varrho, \varphi, z) = 1$, we get an expression for the volume V of domain Ω as a triple integral in cylindrical co-ordinates:

$$V = \iiint_{\Omega} \varrho \, d\varrho \, d\varphi \, dz.$$

Evaluation of a triple integral in cylindrical co-ordinates reduces to single integrations with respect to ϱ, φ and z by using the same

principles as in the case of Cartesian co-ordinates. In particular, if the domain on integration Ω is the interior of the cylinder $\varrho \leqslant r$, $0 \leqslant z \leqslant h$, the limits of the iterated integral are constant and there is no change on changing the order of the integrations:

$$I = \int_0^h dz \int_0^{2\pi} d\varphi \int_0^r F(\varrho, \varphi, z)\, \varrho\, d\varrho.$$

II. SPHERICAL CO-ORDINATES. We now refer the domain of integration Ω to a *system of spherical (polar) co-ordinates* $(\varrho, \varphi, \theta)$ (see Sec. 152). We divide Ω into sub-domains v_i by the three systems of *co-ordinate surfaces:* $\varrho = \text{const.}$, $\varphi = \text{const.}$, $\theta = \text{const.}$, these

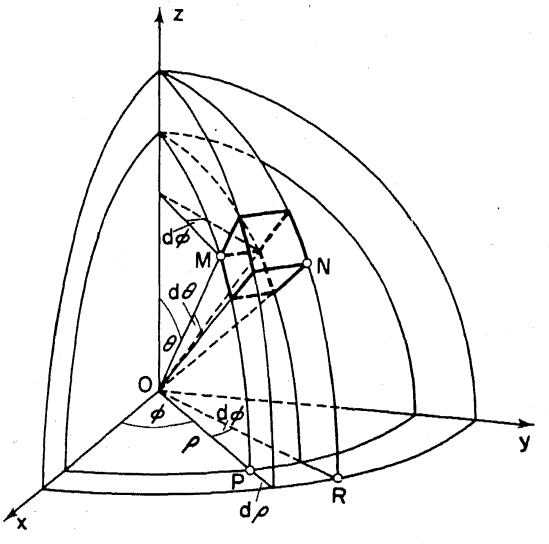

FIG. 47

being respectively spheres with centre at the origin, half-planes through Oz, and cones with vertex at the origin and axes coinciding with one of the semi-axes Oz. Sub-domains v_i consist of "hexahedrons" (Fig. 47). Neglecting higher order infinitesimals, hexahedron MN can be regarded as a rectangular parallelepiped with the following dimensions: $d\varrho$ in the direction of the radius vector, $\varrho\, d\theta$ in the direction of the meridian, $\varrho \sin\theta\, d\varphi$ in the direction of the parallel. We now have for an elementary volume:

$$dv = \varrho^2 \sin\theta\, d\varrho\, d\varphi\, d\theta.$$

Hence

$$I = \iiint_\Omega F(P)\, dv = \iiint_\Omega F(\varrho, \varphi, \theta)\, \varrho^2 \sin\theta\, d\varrho\, d\varphi\, d\theta.$$

When the systems of Cartesian and spherical co-ordinates are arranged as shown in Fig. 47, we have $x = \varrho \sin\theta \cos\varphi$, $y = \varrho \sin\theta \sin\varphi$, $z = \varrho \cos\theta$; the formula for transforming a triple integral from Cartesian to spherical co-ordinates is thus

$$I = \iiint_{\Omega} f(x, y, z)\, dx\, dy\, dz = \iiint_{\Omega} F(\varrho, \varphi, \theta)\, \varrho^2 \sin\theta\, d\varrho\, d\varphi\, d\theta,$$

where

$$F(\varrho, \varphi, \theta) = f(\varrho \sin\theta \cos\varphi,\ \varrho \sin\theta \sin\varphi,\ \varrho \cos\theta).$$

When $f(\varrho, \varphi, \theta) = 1$, we get an expression for the volume V of domain Ω as a triple integral in spherical co-ordinates:

$$V = \iiint_{\Omega} \varrho^2 \sin\theta\, d\varrho\, d\varphi\, d\theta.$$

Evaluation of a triple integral in spherical co-ordinates reduces as in the previous cases to single integrations with respect to ϱ, φ and θ.

If the domain of integration Ω contains the origin and is *star-shaped* with respect to it (i.e. its boundary is cut by any radius vector from the origin in one point), we have

$$I = \iiint_{\Omega} F(\varrho, \varphi, \theta)\, \varrho^2 \sin\theta\, d\varrho\, d\varphi\, d\theta$$

$$= \int_0^{\pi} \sin\theta\, d\theta \int_0^{2\pi} d\varphi \int_0^{\alpha(\varphi, \theta)} F(\varrho, \varphi, \theta)\, \varrho^2\, d\varrho,$$

where $\varrho = \alpha(\varphi, \theta)$ is the spherical equation of the surface bounding domain Ω. In particular, when this surface is a sphere $\varrho = R$, the limits of all three ordinary integrals are constants and there is no change on changing the order of the integrations; we get

$$I = \int_0^{\pi} \sin\theta\, d\theta \int_0^{2\pi} d\varphi \int_0^{R} F(\varrho, \varphi, \theta)\, \varrho^2\, d\varrho.$$

Obviously, evaluation of a triple integral over a domain of this type is particularly simple with spherical co-ordinates.

III. GENERAL CASE OF EVALUATION IN SPHERICAL CO-ORDINATES. SOLID ANGLE. It is worth discussing in detail the evaluation of a triple integral in the general case in spherical co-ordinates.

Let Ω not contain the origin and let its boundary be cut by any radius vector from the origin in not more than two points. We shall find the most convenient method of reducing a triple integral I to an iterated integral.

We circumscribe Ω with a conical surface with vertex at the origin. It touches Ω along a curve L which divides the boundary surface of the domain into two parts; let the equation of the part nearest the origin be $\varrho = \alpha_1(\varphi,\theta)$, and of the other part: $\varrho = \alpha_2(\varphi, \theta)$. Our conical surface "cuts out" a domain K from the surface of the unit sphere (i.e. the sphere with centre at the origin and unit radius), K being the central projection (from the origin) of domain Ω on to the unit sphere. We shall integrate first along the direction of the radius vector. This is done by integrating the integrand

$$F(\varrho, \varphi, \theta)\,\varrho^2 \sin\theta$$

over the segment contained in Ω of the straight line through the origin and a point $P(\varphi, \theta)$ of domain K. For given φ and θ, the variable of integration ϱ will vary from $\alpha_1(\varphi, \theta)$ to $\alpha_2(\varphi, \theta)$; $\alpha_1(\varphi, \theta)$ is the radius vector of the point of "entry" of the line into Ω, and $\alpha_2(\varphi, \theta)$ the radius vector of its point of "departure" from Ω. The result of the integration is a quantity dependent on the point $P(\varphi, \theta)$; we shall denote it by $\Phi(\varphi, \theta)$:

$$\Phi(\varphi, \theta) = \int_{\alpha_1(\varphi, \theta)}^{\alpha_2(\varphi, \theta)} F(\varrho, \varphi, \theta)\,\varrho^2 \sin\theta\, d\varrho.$$

During the integration φ and θ are assumed constant.

We obtain the value of the triple integral if we can find the double integral

$$\iint_K \Phi(\varphi, \theta)\, d\varphi\, d\theta,$$

i.e.

$$I = \iint_K \sin\theta\, d\varphi\, d\theta \int_{\alpha_1(\varphi, \theta)}^{\alpha_1(\varphi, \theta)} F(\varrho, \varphi, \theta)\,\varrho^2\, d\varrho.$$

The expression $\sin\theta\, d\theta\, d\varphi$ may easily be seen (Fig. 47) to give the element dq of area of the surface of the unit sphere. With the aid of this we can write the triple integral in spherical co-ordinates as

$$\iint_K dq \int_{\alpha_1(\varphi, \theta)}^{\alpha_2(\varphi, \theta)} F(\varrho, \varphi, \theta)\,\varrho^2\, d\varrho. \qquad (*)$$

The double integral is said to be taken here over surface K of the unit sphere into which domain Ω is projected.

Definition. The area of the part of the unit sphere representing the central projection (from the origin) of a surface is called the solid angle *of the surface.*

We can therefore say that an elementary volume dv of a body in spherical co-ordinates is the product of $\varrho^2\, d\varrho$ and the solid angle dq of the element of surface bounding the body. This is entirely analogous to the fact that the elementary area $d\sigma$ of a domain in polar co-ordinates is equal to the product of $\varrho\, d\varrho$ and the length of arc (angle) $d\varphi$ of the unit circle into which an element of the curve bounding the domain is centrally projected.

The limits of integration when next integrating with respect to φ are obtained as the values of this co-ordinate for the points of "entry" into and "departure" from K of a parallel with constant but arbitrary θ; these limits in general depend on θ: let $\beta_1(\theta) \leqslant \varphi \leqslant \beta_2(\theta)$. Finally, we integrate with

respect to θ between the limits of maximum variation of θ in the spherical domain K, i.e. $\theta_1 \leqslant \theta \leqslant \theta_2$. We thus obtain a formula for transforming to an iterated integral:

$$I = \iiint_{\Omega} F(\varrho, \varphi, \theta)\, \varrho^2 \sin\theta\, d\varrho\, d\varphi\, d\theta$$
$$= \int_{\theta_1}^{\theta_2} \sin\theta\, d\theta \int_{\beta_1(\theta)}^{\beta_2(\theta)} d\varphi \int_{\alpha_1(\varphi,\theta)}^{\alpha_2(\varphi,\theta)} F(\varrho, \varphi, \theta)\, \varrho^2\, d\varrho.$$

We conclude by observing that formula (*) also applies when the domain is star-shaped with respect to the origin. In this case $\alpha_1 = 0$ and K is the entire unit sphere.

4. Applications of Double and Triple Integrals

176. Approach for the Solution of Problems. We shall indicate the nature and general approach to the solution of problems leading to the evaluation of a double integral. This approach is a natural generalization of the one for solving problems on applications of the ordinary integral (see Sec. 115) and indicates a unique short cut whereby the solution is put in the form of a double integral. We shall follow a similar procedure to that for the ordinary integral and make use of the *Newton–Leibniz formula*, given in Sec. 170.

Suppose a quantity u can be regarded as a function $u = F(\sigma)$ of a variable domain σ. Further, on subdivision of domain σ, let u be the sum of the similar terms corresponding to the subdivisions (property of *additiveness**). Assuming that $F(\sigma)$ is differentiable with respect to the domain, let it be required to find the value of $u = F(D)$ corresponding to some domain D. We have by the Newton–Leibniz formula:

$$F(D) = \iint_{D} dF(\sigma).$$

Hence, given the stated conditions, *the required quantity is measured by the double integral of the differential of the unknown function over the domain in question.*

We can, however, as in the problem of the volume of a cylindrical body, subdivide the given domain D and express $F(D)$ as a sum of terms corres-

* As an aid to following the discussion we recommend the reader to visualize at each step the simple problem of the volume of a cylindrical body. The variable volume of a cylindrical body bounded by a given surface is a function of the variable domain at the base of the cylinder. Expressing the total volume as the sum of the volumes of partial cylindrical bodies corresponds to subdividing the base.

ponding to the subdivisions; on then replacing each such term, which is unknown to us, by a known quantity, we can pass to the limit as the number of subdivisions increases indefinitely and their diameters tend to zero.

All in all, we obtain as the value of the required quantity an integral of the differential of the quantity.

But the theory of the double integral (Newton–Leibniz formula) enables us to avoid all this working when solving a concrete problem of the type in question (i.e. we do not need to repeat the definition of integral); we only need to make use of the final result of the theory, that *the required quantity is equal to the integral of its differential.*

There are thus three stages in the solution of a problem:

I. FORMATION OF THE "DIFFERENTIAL EQUATION", i.e. formation of the expression for the differential of the required quantity u as a function of domain σ, $u = F(\sigma)$, in terms of given functions and the elementary area $d\sigma$ of the domain:

$$du = dF(\sigma) = f(P)\,d\sigma, \qquad (*)$$

where $f(P)$ is a function of a point equal to $F'(\sigma)$.

For this, an infinitesimal domain $d\sigma$ is taken at an arbitrary point P of the domain and the differential $dF(\sigma)$ found from the condition that it is an infinitesimal proportional to $d\sigma$ and equivalent to the part of $F(\sigma)$ corresponding to $d\sigma$. In concrete problems we usually take as $dF(\sigma)$ the quantity that is obtained when the magnitudes defining it preserve (in $d\sigma$) their values at the point P. For instance, the differential at point P of the volume of a cylindrical body is the volume of the cylinder with base $d\sigma$ and with constant z co-ordinate equal to the z co-ordinate of the given surface at P.

II. PASSAGE FROM EQUATION (*) TO THE DOUBLE INTEGRAL

$$F(D) = \iint_D f(P)\,d\sigma.$$

The differential $dF(\sigma) = f(P)\,d\sigma$ gives the "element" of the quantity $F(\sigma)$ (whereas its true part is the part which strictly corresponds to domain $d\sigma$). This passage to the integral is embodied in the "summation of all the elements", leading to the accurate result.

III. EVALUATION OF THE DOUBLE INTEGRAL. We possess adequate means for this, i.e. reduction of the double to an iterated integral, in other words, to a sequence of single integrations.

The above general method of solution can be carried over directly to triple integrals.

We shall illustrate the solution of a problem by this method by taking the problem of *finding the mass of a distribution of material, given its density*, and considering the connection between these two important physical concepts. This problem was discussed for curves in Sec. 42, II, and 85, III. We now consider the problem for a plane domain D. Let mass be distributed continuously in this domain. A definite mass m corresponds to each domain σ in D, so that m is a function of σ, i.e. $m = F(\sigma)$. *The mass is said to be distributed uniformly if any two parts of D of equal area contain equal masses.* The ratio of the mass of a piece of D to its area is a constant in this case, numerically equal to the mass of a unit area situated at any place in the domain. This ratio is called the *surface density of the material* in domain D. If the mass is distributed non-uniformly, however, an accurate definition of density at a given position is needed, i.e. at a given point P of the domain.

Let us take a neighbourhood $\Delta\sigma$ of point P with an area denoted by the same symbol $\Delta\sigma$. On dividing the mass $\Delta m = \Delta F$ contained in $\Delta\sigma$ by the area $\Delta\sigma$, we obtain the mass which would be contained per unit area of the piece $\Delta\sigma$ if it were uniformly distributed. The quantity $\Delta F(\sigma)/\Delta\sigma$ is called the *mean surface density* in the domain $\Delta\sigma$.

Since we want to introduce the concept of density at a point P, we let the diameter of domain $\Delta\sigma$ diminish indefinitely (contract to the point P). The surface density at P is defined as the number to which the mean density tends when the diameter of the corresponding domain $\Delta\sigma$ tends to zero.

In other words, *the surface density δ of the material at point P is given by the derivative of the mass with respect to the area:*

$$\delta = F'(\sigma) = \lim_{\Delta\sigma \to 0} \frac{\Delta F(\sigma)}{\Delta\sigma}.$$

The density is a *function of point P* in domain D:

$$\delta = \mu(P).$$

Let us now start, conversely, from a given density δ in domain D as a function of point P. We have for the differential of the mass, in accordance with the above:

$$dm = dF(\sigma) = \mu(P)d\sigma.$$

(This is the mass in the infinitesimal domain $d\sigma$ on condition that the density is constant in this domain and equal to the density at point P.)

We get by "summing" all the elements of mass in domain D:

$$m = F(D) = \iint_D \mu(P)\,d\sigma.$$

The mass of material in D is measured by the double integral of the surface density over the domain.

The idea of the reciprocity of the physical concepts of mass and density is expressed in the theorem on the connection between the derivative with respect to a domain and the integral over a domain.

It should be noted that "mass" and "density" can refer here not merely to matter but also e.g. to electricity (the "quantity" of electricity and the "charge") and so on. The whole of our discussion of mass and density in a plane domain D can be carried over to a spatial domain Ω merely by changing the terminology and notation.

Example. Let us find the mass of a sphere Ω of radius R if the volume density is given by the function

$$\mu(P) = \frac{1}{\sqrt{R^2 - \varrho^2}},$$

where $\varrho = \sqrt{x^2 + y^2 + z^2}$ is the radius vector of point P.

We have:

$$m = \iiint_\Omega \frac{dv}{\sqrt{R^2 - \varrho^2}}.$$

This integral is conveniently evaluated with the aid of spherical co-ordinates (Sec. 176):

$$m = \int_0^\pi \sin\theta\,d\theta \int_0^{2\pi} d\varphi \int_0^R \frac{\varrho^2\,d\varrho}{\sqrt{R^2 - \varrho^2}} = 4\pi \int_0^R \frac{\varrho^2\,d\varrho}{\sqrt{R^2 - \varrho^2}}.$$

Substituting $\varrho = R\sin t$ gives

$$m = 4\pi R^2 \int_0^{\frac{\pi}{2}} \sin^2 t\,dt = \pi^2 R^2.$$

This solution is readily obtained, however, without recourse to a triple integral, since the surface density on any concentric sphere $\varrho = \text{const}\ (< R)$ is constant by hypothesis, and the total mass is obtained by "summation" over a radius (i.e. by a single integration) of the masses on the concentric spheres with $0 \leqslant \varrho \leqslant R$.

177. Some Geometrical Problems. We first of all recall the problem of finding the volume of a body. Since we have continually illustrated the theory of double integrals by reference to this problem there is no need to dwell on it further. We shall merely mention that certain problems do not require the use of double integrals, since the result is more readily obtained by means of an ordinary integral. This happens, for instance, when the areas of parallel sections of the body are already known, and in particular, when we are dealing with a solid of revolution (see Sec. 119).

We return to the problem of finding the *area of a surface.* Let the surface be given by the equation $z = f(x, y)$, where $f(x, y)$ is a differentiable function with continuous partial derivatives, and let the area Q be required of the part K corresponding to a domain D in the Oxy plane, i.e. D is the projection of K.

The solution is based on two principles as in previous similar cases (Secs. 118, 123), viz.

(1) the surface area has the property of additiveness;

(2) the differential of the surface area is equal to the area of the piece of tangent plane corresponding to an infinitesimal piece of the surface*.

We find an expression for the differential dq of the area with the aid of these principles, then an expression for the whole of area Q with the aid of the integral.

We take a point $P_0(x_0, y_0)$ in domain D and its neighbourhood $d\sigma$ ($d\sigma$ denotes both the neighbourhood and its area). We draw the tangent plane T (Fig. 48) at the point of the surface $M_0(x_0, y_0, z_0 = f(x_0, y_0))$:

$$z - z_0 = f'_x(x_0, y_0)\,(x - x_0) + f'_y(x_0, y_0)\,(y - y_0).$$

The area of the piece of T corresponding to domain $d\sigma$ is in fact the required differential dq; since $d\sigma$ is the orthogonal projection of the piece of T on Oxy, we have

$$d\sigma = dq \cos\gamma,$$

where γ is the angle between planes T and Oxy. This angle is equal to the angle between the normal to plane T and Oz. We thus have for $\cos\gamma$ (see Sec. 166):

$$\cos\gamma = \frac{1}{\sqrt{1 + f'^2_x(x_0, y_0) + f'^2_y(x_0, y_0)}},$$

i.e.

$$dq = \sqrt{1 + f'^2_x(x_0, y_0) + f'^2_y(x_0, y_0)}\, d\sigma.$$

* We recall the similar correspondence of the differential of the arc of a curve to the length of the segment of the tangent line in question.

This is the *expression for the elementary area (or simply element) of the surface* $z = f(x, y)$ *at the point* $M_0(x_0, y_0, z_0)$.

On "summing" all the surface elements over domain D, we obtain for the required area Q:

$$Q = \iint_D \sqrt{1 + z_x'^2 + z_y'^2}\, d\sigma = \iint_D \sqrt{1 + z_x'^2 + z_y'^2}\, dx\, dy. \qquad (*)$$

We recommend the reader to deduce from (*) the formula of Sec. 120 for the area of a surface of revolution.

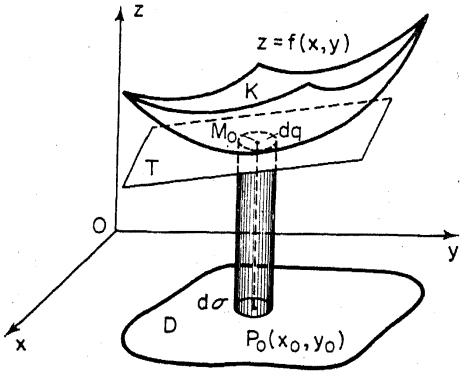

FIG. 48

A surface which has an area in accordance with this definition is said to be integrable.

Example 1. Let us find the area of a sphere of radius R. The sphere is given by $x^2 + y^2 + z^2 = R^2$. We have:

$$z_x' = -\frac{x}{\sqrt{R^2 - x^2 - y^2}}, \quad z_y' = -\frac{y}{\sqrt{R^2 - x^2 - y^2}}.$$

Hence we find for the hemisphere:

$$\frac{1}{2}Q = \iint_D \frac{R}{\sqrt{R^2 - x^2 - y^2}}\, dx\, dy,$$

where D is the circle $x^2 + y^2 \leqslant R^2$. We find on passing to polar co-ordinates:

$$\frac{1}{2}Q = R \iint_D \frac{\varrho\, d\varrho\, d\varphi}{\sqrt{R^2 - \varrho^2}} = R \int_0^{2\pi} d\varphi \int_0^R \frac{\varrho\, d\varrho}{\sqrt{R^2 - \varrho^2}},$$

i.e.

$$\frac{1}{2}Q = 2\pi R^2, \quad \text{whence } Q = 4\pi R^2.$$

Example 2. We shall take as our second example a problem on *the mass distributed on a surface.*

The concept of surface density at a point of a surface is set up precisely as in the case of curves, plane domains and spatial domains. The density δ at a point M of a surface is defined as the limit of the ratio of the mass Δm of a part of the surface containing point M to the area Δq of the part when its diameter tends to zero:

$$\delta = \lim_{\Delta q \to 0} \frac{\Delta m}{\Delta q} = \frac{dm}{dq},$$

i.e. *the density is the derivative of the mass with respect to the surface.* The density is a function of the point M of the surface: $\delta = \mu(M)$.

Conversely, if the density is given as a function of point M of the surface: $\delta = \mu(M)$, we have

$$dm = \mu(M)\, dq,$$

and therefore

$$m = \iint_K \mu(M)\, dq,$$

where K is the part of the surface on which the mass is evaluated.

178. Some Problems of Statics.

I. PLANE LAMINA. We shall find with the aid of a double integral the centre of gravity of a plane domain D on which mass is distributed (a material plane lamina). The mass distribution is characterized by the density $\delta = \mu(P)$ (see Sec. 121).

Let us evaluate the statical moments M_x and M_y of the lamina about axes Ox and Oy (in its plane).

Let $d\sigma$ be the infinitesimal neighbourhood of any point $P(x, y)$. Its mass is $\mu(P)\, d\sigma$ to an accuracy of higher order infinitesimals. The element dm_x of the statical moment about Ox is found by assuming the mass of $d\sigma$ to be concentrated at point P and taking the moment of P about this axis, i.e.

$$dm_x = y\mu(P)\, d\sigma.$$

Similarly

$$dm_y = x\mu(P)\, d\sigma.$$

Integration over D gives us

$$M_x = \iint_D y\mu(P)\, d\sigma, \quad M_y = \iint_D x\mu(P)\, d\sigma.$$

We have further, by definition of the centre of gravity,

$$\xi M = M_y, \quad \eta M = M_x,$$

where ξ, η are the co-ordinates of the centre of gravity and M is the mass of the lamina, equal to

$$\iint_D \mu(P)\, d\sigma.$$

Hence

$$\xi = \frac{\iint_D x\mu(P)\, d\sigma}{\iint_D \mu(P)\, d\sigma}, \quad \eta = \frac{\iint_D y\mu(P)\, d\sigma}{\iint_D \mu(P)\, d\sigma}.$$

In the case of a homogeneous lamina ($\mu(P) = \text{const}$) the formulae can be simplified by cancelling μ.

If D is a curvilinear trapezium with base say on Ox, direct integration with respect to y and x in these formulae leads to the expressions found in Sec. 121.

We now consider centres of gravity in space.

Definition. The centre of gravity of a system *of material particles in space is the point such that, if the total mass of the system were concentrated at this point, the statical moments about the co-ordinate planes would be equal to the corresponding statical moments of the system* (cf. Sec. 121). (The statical moment of a particle about a plane is equal to the mass of the particle multiplied by its distance from the plane.)

Finding the co-ordinates of the centre of gravity thus reduces to finding the statical moments of the system and its mass.

The co-ordinates of the centre of gravity of a continuous distribution are found with the aid of integration, on the basis of the fact that the statical moment of the whole is equal to the sum of the moments of the component parts (property of additiveness).

II. Solids. The centre of gravity of a spatial domain Ω in which mass is distributed with a volume density $\delta = \mu(P)$ (a material body) is found in the same way as above.

We evaluate the statical moments M_{xy}, M_{xz}, M_{yz} of the solid about the co-ordinate planes Oxy, Oxz, Oyz.

Let dv be an infinitesimal neighbourhood of point $P(x, y, z)$. Its mass is $\mu(P)\, dv$, to an accuracy of higher order infinitesimals. Element dm_{xy} of the statical moment with respect to Oxy is obtained by assuming that the mass of dv is concentrated at P and taking the statical moment about Oxy. We have

$$dm_{xy} = z\mu(P)\, dv.$$

Similarly,

$$dm_{xz} = y\mu(P)\, dv, \quad dm_{yz} = x\mu(P)\, dv.$$

Integration over domain Ω gives

$$M_{xy} = \iiint_\Omega z\mu(P)\, dv, \quad M_{xz} = \iiint_\Omega y\mu(P)\, dv, \quad M_{yz} = \iiint_\Omega x\mu(P)\, dv.$$

We have by definition of the centre of gravity:

$$\xi M = M_{yz}, \quad \eta M = M_{xz}, \quad \zeta M = M_{xy},$$

where ξ, η, ζ are the co-ordinates of the centre of gravity and M is the mass of the solid. Hence

$$\xi = \frac{\iiint\limits_{\Omega} x\mu(P)\,dv}{\iiint\limits_{\Omega} \mu(P)\,dv}, \quad \eta = \frac{\iiint\limits_{\Omega} y\mu(P)\,dv}{\iiint\limits_{\Omega} \mu(P)\,dv}, \quad \zeta = \frac{\iiint\limits_{\Omega} z\mu(P)\,dv}{\iiint\limits_{\Omega} \mu(P)\,dv}.$$

If Ω is a solid of revolution (about one of the co-ordinate axes), the expression of Sec. 121 is readily obtained from these formulae for the co-ordinate in question; the other two co-ordinates are zero.

Example. Let us find the centre of gravity of the homogeneous hemisphere Ω:

$$x^2 + y^2 + z^2 \leqslant R^2, \quad z \geqslant 0,$$

with density $\delta = \text{const}$.

We have for moment M_{xy}:

$$M_{xy} = \delta \iiint\limits_{\Omega} z\,dv.$$

The integral is readily evaluated on passing to spherical co-ordinates:

$$M_{xy} = \delta \iiint\limits_{\Omega} \varrho\cos\theta\,\varrho^2\sin\theta\,d\varrho\,d\varphi\,d\theta = \delta \int_0^{\frac{\pi}{2}} \sin\theta\cos\theta\,d\theta \int_0^{2\pi} d\varphi \int_0^{R} \varrho^3\,d\varrho$$

$$= \delta \cdot \frac{1}{2} \cdot 2\pi \cdot \frac{R^4}{4} = \frac{1}{4}\delta\pi R^4.$$

Since the mass of Ω is $^2/_3\,\delta\pi R^3$, we have

$$\zeta = \frac{M_{xy}}{M} = \frac{\frac{1}{4}\delta\pi R^4}{\frac{2}{3}\delta\pi R^3} = \frac{3}{8}R.$$

As regards the other two co-ordinates of the centre of gravity (ξ and η), direct working shows that they are zero; this is also immediately obvious, since the hemisphere is symmetrical with respect to Oz (a solid of revolution with axis Oz).

III. CURVED LAMINA. We can find in precisely the same way the centre of gravity of part K of a surface on which mass is distributed with surface density $\delta = \mu(P)$ (material curved lamina).

We shall omit the arguments, which have been given twice, and write down directly the expressions for the statical moments M_{xy}, M_{xz}, M_{yz} of lamina K about the co-ordinate planes:

$$M_{xy} = \iint\limits_{K} z\mu(M)\,dq, \quad M_{xz} = \iint\limits_{K} y\mu(M)\,dq, \quad M_{yz} = \iint\limits_{K} x\mu(M)\,dq,$$

where dq is an element of the surface.

We find for the co-ordinates of the centre of gravity:

$$\xi = \frac{\iint\limits_K x\mu(M)\,dq}{\iint\limits_K \mu(M)\,dq}, \quad \eta = \frac{\iint\limits_K y\mu(M)\,dq}{\iint\limits_K \mu(M)\,dq}, \quad \zeta = \frac{\iint\limits_K z\mu(M)\,dq}{\iint\limits_K \mu(M)\,dq}.$$

If K is a surface of revolution (about one of the axes), the above formulae yield the expression of Sec. 121 for the co-ordinate concerned; the other two co-ordinates are zero.

Example. Let us find the centre of gravity of the homogeneous hemispherical shell K:

$$x^2 + y^2 + z^2 = R^2,\ z \geqslant 0,$$

with density $\delta = \text{const}$.
We have for the moment M_{xy}:

$$M_{xy} = \delta \iint\limits_K z\,dq.$$

In view of the fact that

$$z = \sqrt{R^2 - x^2 - y^2}, \quad \text{and} \quad dq = \frac{R}{\sqrt{R^2 - x^2 - y^2}}\,dx\,dy$$

(see Sec. 177), we have

$$M_{xy} = \delta \iint\limits_D R\,dx\,dy = \delta R\,\pi R^2 = \delta\,\pi R^3.$$

Since the mass is $2\,\delta\,\pi R^2$, we have

$$\zeta = \frac{M_{xy}}{M} = \frac{\delta\,\pi R^3}{2\,\delta\,\pi R^2} = \frac{1}{2}R.$$

The other co-ordinates (ξ and η) are obviously zero.

The construction of the formulae for the co-ordinates of the centre of gravity is precisely the same in all the above cases; any of them is easily written down once the general principle is mastered.

It may be mentioned in conclusion that the co-ordinates of the centre of gravity for a homogeneous mass distribution depend only on the form and disposition of the solid and not on its density.

IV. MOMENTS OF INERTIA. *The moment of inertia of a particle $M(x, y, z)$ of mass m with respect to a plane (say Oyz), an axis (say Oz), or a point (say the origin), is the product of mass m and the square of the distance of M from the plane (mx^2), from the axis ($m(x^2 + y^2)$) or from the point ($m(x^2 + y^2 + z^2)$).*

The passage from the moment of inertia of a particle to that of a continuous distribution (plane or curved lamina, solid) is similar to the corresponding passage for statical moments. Obviously, the expressions for the moment of inertia of a plane lamina about a co-ordinate axis or for the moments of inertia of a surface or solid with respect to a co-ordinate plane only differ from the corresponding expressions for the statical moments (see I, II, III) in having the squares instead of the first powers of the co-ordinates under the integrals.

As regards the moments of inertia about the axes and origin in the three-dimensional case, the derivation of the expressions follows the same lines as when applying the concept of integral, and the discussion is the same as in all the above problems. For instance, we find for the moment of inertia I_z of a solid with respect to Oz:

$$I_z = \iiint_{\Omega} (x^2 + y^2)\, \mu(P)\, dv,$$

where $\mu(P)$ is the density of the solid at P.

It may be mentioned that the significance of the moment of inertia of a solid about an axis lies in the close connection between this and the kinetic energy of the solid when it rotates about the axis. Let the solid Ω rotate about Oz with constant angular velocity ω. Let us find the kinetic energy J_z of the solid. We know that the kinetic energy of a particle is given by $\frac{1}{2}mv^2$, where m is the mass of the particle and v is its velocity. The kinetic energy of a system of particles is defined as the sum of the kinetic energies of the individual particles, whilst the kinetic energy of a solid is the sum of the kinetic energies of all its component parts. This fact enables us to apply the concept of integral for evaluating the kinetic energy.

We take a neighbourhood dv of a point $P(x, y, z)$ of solid Ω. The linear velocity v of point P on rotation about Oz is $\omega\sqrt{x^2 + y^2}$, i.e. the kinetic energy of the part dv of Ω is, to an accuracy of higher order infinitesimals:

$$\frac{1}{2}\mu(P)\, dv \,.\, \omega^2(x^2 + y^2),$$

where $\mu(P)$ is the density at P. We obtain for the kinetic energy of the whole of Ω:

$$J_z = \iiint_{\Omega} \frac{1}{2}\omega^2(x^2 + y^2)\, \mu(P)\, dv = \frac{1}{2}\omega^2 \iiint_{\Omega} (x^2 + y^2)\, \mu(P)\, dv,$$

i.e.

$$J_z = \frac{1}{2}\omega^2 I_z.$$

The kinetic energy of a solid rotating about an axis with constant angular velocity is equal to half the square of the angular velocity times the moment of inertia of the solid about the axis of rotation.

5. Improper Integrals. Integrals Dependent on a Parameter

179. Improper Double and Triple Integrals. Just as the concept of ordinary integral may be extended to the cases of an infinite interval of integration and discontinuous integrands (see Secs. 110, 113), the concepts of double and triple integrals may likewise be extended to an infinite domain of integration and discontinuous integrands.

I. INTEGRALS OVER INFINITE DOMAINS. Let the function $z = f(x, y)$ be continuous in a domain D stretching to infinity.

We consider the double integral

$$I(B) = \iint_B f(x, y)\, dx\, dy,$$

over a finite domain B lying wholly in D. We expand B in some arbitrary manner so as to contain and retain any given point of domain D, this being written as $B \to D$ (domain B tends to D).

Definition. **The *improper double integral* of a function $f(x, y)$ over the domain D is the limit I (if it exists) of integral $I(B)$ as $B \to D$.**

In other words,

$$\iint_D f(x, y)\, dx\, dy = \lim_{B \to D} \iint_B f(x, y)\, dx\, dy.$$

We can also say that *the improper integral on the left-hand side exists*, or *is convergent*. If the limit of $I(B)$ does not exist or is infinite, we say that *the improper integral does not exist*, or *is divergent*.

The existence of an improper double integral means from the geometrical point of view that the corresponding cylindrical body with infinite base can be assigned a definite volume.

It should be mentioned that the infinite domain D over which the improper integral is taken can be the whole of Oxy.

Example 1. We take the integral

$$\iint_D e^{-x^2-y^2}\, dx\, dy,$$

over the whole of Oxy. We shall prove the existence of this improper double integral and find its value.

We take the (proper) integral of the integrand over a circular disc B_ν of radius ν with centre at the origin:

$$I(B_\nu) = \iint_{x^2+y^2 \leqslant \nu^2} e^{-x^2-y^2} dx\, dy$$

This is evaluated by passing to polar co-ordinates. We have

$$I(B_\nu) = \iint_{B_\nu} e^{-\varrho^2}\, \varrho\, d\varrho\, d\varphi = \int_0^{2\pi} d\varphi \int_0^{\nu} e^{-\varrho^2}\, \varrho\, d\varrho = \pi(1 - e^{-\nu^2}).$$

We have as $\nu \to \infty$:

$$\lim_{\nu \to \infty} I(B_\nu) = \lim_{\nu \to \infty} \pi\,(1 - e^{-\nu^2}) = \pi.$$

We now take the (proper) integral over an arbitrary domain B containing the origin:

$$I(B) = \iint_B e^{-x^2-y^2}\,dx\,dy.$$

Since $e^{-x^2-y^2} \geqslant 0$, we have

$$I(B_r) \leqslant I(B) \leqslant I(B_R),$$

where B_r denotes a circular domain inscribed in domain B and B_R a circular domain circumscribing B. When B expands so as to

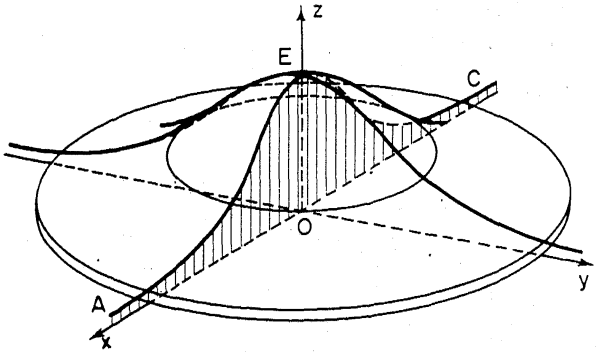

FIG. 49

fill the entire Oxy plane, $r \to \infty$ and $R \to \infty$; therefore, by what has been proved, $I(B_r) \to \pi$, $I(B_R) \to \pi$, i.e.

$$I(B) \to \pi$$

on arbitrary expansion of B over the entire plane.

Thus

$$\iint_D e^{-x^2-y^2}\,dx\,dy = \pi.$$

This integral gives the volume of the infinite solid bounded by Oxy and the surface formed by rotation of the curve $z = e^{-x^2}$ (in the Oxz plane) about Oz (Fig. 49).

Let us take as domain B the square: $-a \leqslant x \leqslant a$, $-a \leqslant y \leqslant a$ $(a > 0)$. We get

$$\iint_B e^{-x^2-y^2}\,dx\,dy = \int_{-a}^{a} e^{-x^2}\,dx \int_{-a}^{a} e^{-y^2}\,dy = \left(\int_{-a}^{a} e^{-x^2}\,dx\right)^2.$$

We find on passing to the limit as $B \to D$, i.e. as $a \to \infty$:

$$\left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)^2 = \iint_D e^{-x^2-y^2}\,dx\,dy = \pi,$$

whence

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}\,.$$

We have thus found the value of Poisson's integral (see Sec. 110). This integral gives the area of any section of the solid illustrated in Fig. 49 by a plane through Oz, say the area of section AEC.

Example 2. It may easily be shown by transforming to polar co-ordinates that the improper integral

$$\int\int \frac{dx\,dy}{(\sqrt{x^2+y^2})^m},$$

over the whole of the plane except the neighbourhood of the origin (0, 0) exists or not, depending on whether $m > 2$ or $m \leqslant 2$.

The concept of triple integral is generalized for infinite spatial domains in precisely the same way.

For instance, it is readily seen by passing to spherical co-ordinates that the improper integral

$$\int\int\int \frac{dx\,dy\,dz}{(\sqrt{x^2+y^2+z^2})^m},$$

over the whole of space except for the neighbourhood of (0, 0, 0) either exists or not, depending on whether $m > 3$ or $m \leqslant 3$.

II. INTEGRALS OF DISCONTINUOUS FUNCTIONS. Let $z = f(x, y)$ be continuous in a finite domain D except for a finite number of individual curves and individual points, at which it has finite jumps. *The integral of $f(x, y)$ over the whole of D is taken to be the sum of the (proper) integrals over the sub-domains into which D can be divided such that the curves and points of discontinuity belong to the boundaries of the sub-domains* (cf. Sec. 112). Therefore, if a finite number of curves and points is removed from the domain of integration, the value of the double integral of a continuous function is unchanged: its value over the remainder of the domain is equal to its value over the whole domain.

Now let $z = f(x, y)$ be continuous at every point of a finite domain D except for a point $P_0(x_0, y_0)$ at which it has an infinite jump.

We consider the double integral

$$I(B) = \int\int_B f(x, y)\,dx\,dy,$$

the domain of integration B being got by removing from D an arbitrary domain containing P_0 and lying in D. We contract the subtracted domain in an arbitrary manner so that its diameter tends to zero; domain B now tends indefinitely to domain D (excluding the point P_0).

Definition. **The *improper double integral* over the domain D of *a discontinuous function* $f(x, y)$ is the limit I (if it exists) of integral $I(B)$ as $B \to D$, i.e.**

$$\iint_D f(x, y)\, dx\, dy = \lim_{B \to D} \iint_B f(x, y)\, dx\, dy.$$

We can also say that the *improper integral* on the left-hand side *exists* or is *convergent*. If $I(B)$ does not tend to a limit or tends to infinity, we say that the *improper integral does not exist* or is *divergent*.

The existence of an improper double integral of a discontinuous function means from the geometrical point of view that a definite volume can be assigned to the corresponding cylindrical body with an infinite "needle".

Example 1. We take the integral

$$\iint_D \ln \frac{1}{\sqrt{x^2 + y^2}}\, dx\, dy,$$

over the circular region $x^2 + y^2 \leqslant 1$, the integrand being discontinuous at the origin. We show that this improper integral exists, and find its value.

We remove the circular piece $x^2 + y^2 \leqslant \nu^2$ $(\nu < 1)$ from D, and take the (proper) integral over the remaining domain B_ν:

$$I(B_\nu) = \iint_{B_\nu} \ln \frac{1}{\sqrt{x^2 + y^2}}\, dx\, dy.$$

This is evaluated by passing to polar co-ordinates:

$$I(B_\nu) = -\iint_{B_\nu} \ln \varrho \,.\, \varrho\, d\varrho\, d\varphi = -\int_0^{2\pi} d\varphi \int_\nu^1 \varrho \ln \varrho\, d\varrho$$

$$= 2\pi \left(\frac{1}{4} + \frac{1}{2} \nu^2 \ln \nu - \frac{1}{4} \nu^2 \right).$$

We have as $\nu \to 0$:

$$\lim_{\nu \to 0} I(B_\nu) = \lim_{\nu \to 0} \left[2\pi \left(\frac{1}{4} + \frac{1}{2} \nu^2 \ln \nu - \frac{1}{4} \nu^2 \right) \right] = \frac{\pi}{2}.$$

As in I, it is easily shown that the integral $I(B)$ tends to the same limit if taken over domain D less the neighbourhood of the origin when the diameter of the neighbourhood tends to zero:

$$\lim_{B \to D} I(B) = \frac{\pi}{2}.$$

Thus

$$\iint_D \ln \frac{1}{\sqrt{x^2 + y^2}}\, dx\, dy = \frac{\pi}{2},$$

which is what we wanted to show.

This integral gives the volume of the cylindrical solid whose base is the circle $x^2 + y^2 \leqslant 1$ and which is bounded by the surface got by revolution of the curve $z = \ln 1/x$ (in Oxz) about Oz (Fig. 50).

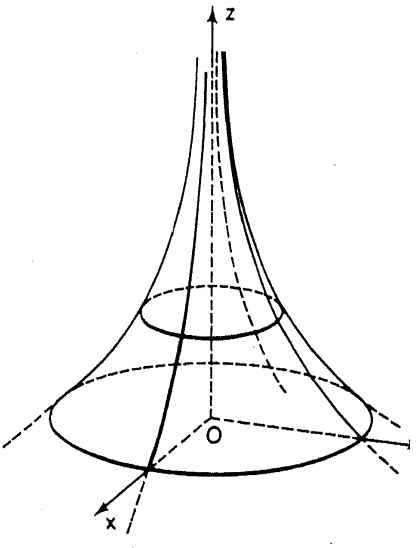

FIG. 50

Example 2. Transformation to polar co-ordinates with pole at the point $P(x, y)$ shows that the improper integral

$$\iint \frac{d\xi\, d\eta}{(\sqrt{(\xi - x)^2 + (\eta - y)^2})^m},$$

over a finite domain containing point $P(x, y)$ either exists or not depending on whether $m < 2$ or $m \geqslant 2$.

The concept of triple integral is similarly generalized to functions of three independent variables having separate surfaces, curves and points of finite discontinuity, or points of infinite discontinuity in the domain of integration.

For instance, it may be seen by passing to spherical co-ordinates in space with pole at the point $P(x, y, z)$ that the improper integral

$$\iiint \frac{d\xi\, d\eta\, d\zeta}{(\sqrt{(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2})^m},$$

over a finite domain in space containing point $P(x, y, z)$, either exists or not, depending on whether $m < 3$ or $m \geqslant 3$.

Improper double and triple integrals naturally have all such properties of proper integrals as are preserved during the passage to the limit.

180. Integrals Dependent on a Parameter. Leibniz's Rule.

Definition. A variable which is independent of the variable of integration but is contained in the integrand or limits of an integral is termed a parameter *of the integral.*

We have several times encountered functions expressible as integrals that depend on a parameter. The most obvious example is the integral with variable upper limit $\int_a^x f(t)\,dt$, which is a function of x. Similarly, we have had inner integrals dependent on one or more parameters when dealing with multiple integration.

The properties of such functions are of great importance.

I. CONTINUITY OF AN INTEGRAL AS A FUNCTION OF A PARAMETER.

THEOREM. **If the function $f(x, \alpha)$ is continuous in x in the interval $[a, b]$, and in α in the interval $[\alpha_1, \alpha_2]$, the integral**

$$F(\alpha) = \int_a^b f(x, \alpha)\,dx$$

is a function continuous in α in the interval $[\alpha_1, \alpha_2]$.

Proof. Let $\alpha \to \alpha_0$, $\alpha_1 \leqslant \alpha_0 \leqslant \alpha_2$. We have:

$$F(\alpha_0) = \int_a^b f(x, \alpha_0)\,dx \qquad (*)$$

and

$$F(\alpha) - F(\alpha_0) = \int_a^b [f(x, \alpha) - f(x, \alpha_0)]\,dx.$$

In view of the continuity of $f(x, \alpha)$, given any positive ε a positive δ can be found such that, when $|\alpha - \alpha_0| < \delta$,

$$|f(x, \alpha) - f(x, \alpha_0)| < \varepsilon_1, \quad \varepsilon_1 = \frac{\varepsilon}{b - a},$$

where, in view of the uniform continuity of $f(x, \alpha)$ (see Sec. 142), δ can be chosen so that the inequality holds for any x, $a \leqslant x \leqslant b$.

The theorem on the upper bound of an integral gives:

$$|F(\alpha) - F(\alpha_0)| \leqslant \varepsilon_1(b - a) = \frac{\varepsilon}{(b - a)}(b - a) = \varepsilon.$$

Hence it follows that

$$\lim_{\alpha \to \alpha_0} F(\alpha) = F(\alpha_0),$$

i.e. $F(\alpha)$ is a continuous function of α in the interval $[\alpha_1, \alpha_2]$. This is what we had to prove.

The result may be written as

$$\lim_{\alpha \to \alpha_0} \int_a^b f(x, \alpha)\, dx = \int_a^b f(x, \alpha_0)\, dx = \int_a^b \lim_{\alpha \to \alpha_0} f(x, \alpha)\, dx.$$

Thus, *if the integrand is continuous, the symbol for the limit and the symbol for the integral can be interchanged.*

II. DIFFERENTIATION OF AN INTEGRAL WITH CONSTANT LIMITS WITH RESPECT TO A PARAMETER. We naturally pose next the question of the differentiability of the function $F(\alpha)$. We assume, in addition to the properties of $f(x, \alpha)$ already mentioned, that it has a continuous partial derivative with respect to α.

THEOREM. **The derivative with respect to a parameter of an integral with constant limits is equal to the integral of the derivative of the integrand with respect to the parameter.**

Proof. Let $\alpha - \alpha_0 = h$. We now have from equation (*), on replacing α_0 by α and α by $\alpha + h$ and applying Lagrange's formula:

$$F(\alpha + h) - F(\alpha) = \int_a^b [f(x, \alpha + h) - f(x, \alpha)]\, dx$$
$$= h \int_a^b f'_\alpha(x, \alpha + \theta h)\, dx, \quad 0 < \theta < 1,$$

whence

$$\frac{F(\alpha + h) - F(\alpha)}{h} = \int_a^b f'_\alpha(x, \alpha + \theta h)\, dx.$$

We pass to the limit as $h \to 0$; bearing in mind the assumed continuity of f'_α, the above rule on interchanging the symbols of limit and integral gives

$$\lim_{h \to 0} \frac{F(\alpha + h) - F(\alpha)}{h} = \int_a^b \lim_{h \to 0} f'_\alpha(x, \alpha + \theta h)\, dx,$$

i.e.

$$F(\alpha) = \int_a^b f'_\alpha(x, \alpha)\, dx.$$

This is what we had to prove.

This proposition regarding the possibility of interchanging the operations of integration and differentiation is known as *Leibniz's rule.*

It should be mentioned that Leibniz's rule also relates to integrals whose limits are variable but independent of the parameter, as also to double and triple integrals, provided the domain of integration is independent of the parameter with respect to which the differentiation is carried out.

It is not in general permissible to apply Leibniz's rule to *improper* integrals depending on a parameter without an extra condition. This amounts to requiring so-called *uniform convergence* of the improper integral.

Definition. An integral depending on a parameter α, $\alpha_1 \leqslant \alpha \leqslant \alpha_2$,

$$I(\alpha) = \int_{a}^{\infty} f(x, \alpha)\, dx$$

is uniformly convergent with respect to α *if, given any positive* ε, *a positive* N *can be found such that, for all* $\eta > N$ *and all* α, $\alpha_1 \leqslant \alpha \leqslant \alpha_2$,

$$\left| \int_{\alpha}^{\infty} f(x, \alpha)\, dx - \int_{\alpha}^{\eta} f(x, \alpha)\, dx \right| = \left| \int_{\eta}^{\infty} f(x, \alpha)\, dx \right| < \varepsilon.$$

It can be shown* that, *if the improper integral* $I(\alpha)$ *is uniformly convergent, it is expressible as a continuous function of* α *in the interval* $[\alpha_1, \alpha_2]$.

Similarly, Leibniz's rule holds for improper integrals, i.e.

$$\frac{d}{d\alpha} \int_{\alpha}^{\infty} f(x, \alpha)\, dx = \int_{\alpha}^{\infty} f'_{\alpha}(x, \alpha)\, dx, \quad \alpha_1 \leqslant \alpha \leqslant \alpha_2,$$

if the improper integral on the right is uniformly convergent.

An example will be given of the use of this rule for evaluating an improper integral.

We take the integral

$$I(\alpha, \beta) = \int_{0}^{\infty} e^{-\alpha x} \frac{\sin \beta x}{x}\, dx,$$

depending on two parameters α and β, where $\alpha > 0$. This integral is uniformly convergent with respect to α and β, since for any β:

$$\left| \int_{\eta}^{\infty} e^{-\alpha x} \frac{\sin \beta x}{x}\, dx \right| < \int_{\eta}^{\infty} \frac{e^{-\alpha x}}{\eta}\, dx = \frac{1}{\alpha \eta e^{\alpha \eta}},$$

whilst the last expression can be made less than any previously assigned ε, $\varepsilon > 0$, for all $\eta > N$ and any $\alpha \geqslant \alpha_1 > 0$. By Leibniz's rule,

$$I'_{\beta}(\alpha, \beta) = \int_{0}^{\infty} e^{-\alpha x} \cos \beta x\, dx. \qquad (**)$$

* See e.g. G.M. Fikhtengol'ts, *Course of differential and integral calculus* (*Kurs differential'nogo i integral'nogo ischisleniya*), vol. II, pp. 730–731, Gost., 1948.

It should be noted that this integral is in fact uniformly convergent, since

$$\left|\int_{\eta}^{\infty} e^{-\alpha x}\cos\beta x\,dx\right| < \int_{\eta}^{\infty} e^{-\alpha x}\,dx = \frac{1}{\alpha e^{\alpha\eta}};$$

given $\alpha \geqslant \alpha_1 > 0$, this expression can be made less than ε, $\varepsilon > 0$, for all $\eta > N_1$ and suitable choice of N_1. However, improper integral (**) is readily evaluated:

$$I'_{\beta}(\alpha,\beta) = e^{-\alpha x}\frac{-\alpha\cos\beta x + \beta\sin\beta x}{\alpha^2+\beta^2}\Bigg|_0^{\infty} = \frac{\alpha}{\alpha^2+\beta^2}.$$

On returning to the function $I(\alpha,\beta)$ we find by indefinite integration with respect to β:

$$I(\alpha,\beta) = \int \frac{\alpha}{\alpha^2+\beta^2}\,d\beta = \arctan\frac{\beta}{\alpha} + C,$$

where C is independent of β. Since the integrand in the original integral vanishes with $\beta = 0$, we have $I(\alpha, 0) = 0$, so that $C = 0$. Hence we have

$$I(\alpha,\beta) = \arctan\frac{\beta}{\alpha}.$$

We now let α tend to zero whilst remaining positive. The argument of the arctan now tends to infinity, with the same sign as β. It can be shown that the integral $I(0,\beta) = \int_0^{\infty}[(\sin\beta x)/x]\,dx$ exists in this case and is equal to $\lim I(\alpha,\beta)$. Therefore $I(0,\beta)$ is equal to $\frac{1}{2}\pi$ or $-\frac{1}{2}\pi$, depending on whether $\beta > 0$ or $\beta < 0$ ($I(0,0) = 0$).

Thus

$$I(\beta) = \int_0^{\infty}\frac{\sin\beta x}{x}\,dx = \begin{cases} \frac{\pi}{2} & \text{for} \quad \beta > 0, \\ 0 & \text{for} \quad \beta = 0, \\ -\frac{\pi}{2} & \text{for} \quad \beta < 0. \end{cases}$$

In particular, we obtain with $\beta = 1$ the value of the Dirichlet integral previously found in Sec. 110.

The function $I(\beta)$ is a remarkable example of a discontinuous function of β given with the aid of a single analytic formula.

This function is known as *Dirichlet's discontinuous multiplier.*

III. DIFFERENTIATION WITH RESPECT TO A PARAMETER OF AN INTEGRAL WITH VARIABLE LIMITS. We now suppose that, given the previous conditions as regards the integrand, the limits of integration are now dependent on α instead of being constant:

$$F(\alpha) = \int_{a(\alpha)}^{b(\alpha)} f(x,\alpha)\,dx. \qquad (***)$$

It is immediately evident that, if functions $a(\alpha)$ and $b(\alpha)$ are continuous in the interval $[\alpha_1, \alpha_2]$, the function $F(\alpha)$ will also be continuous in this interval*.

THEOREM. **The derivative with respect to the parameter of integral(***) with limits dependent on the parameter is equal to**

$$F'(\alpha) = \frac{d}{d\alpha}\int_{a(\alpha)}^{b(\alpha)} f(x, \alpha)\, dx$$

$$= \int_{a(\alpha)}^{b(\alpha)} f'_\alpha(x, \alpha)\, dx + f[b(\alpha), \alpha]\, b'(\alpha) - f[a(\alpha), \alpha]\, a'(\alpha).$$

Proof. The function F depends on α both because of the dependence of the integrand on α and because the limits a and b are functions of α.

We shall write $\varphi(\alpha, a, b)$ for the integral, thus clearly indicating the arguments via which parameter α appears in the function F. We have by the rule for differentiation of a function of a function:

$$F'(\alpha) = \frac{\partial \varphi}{\partial \alpha} + \frac{\partial \varphi}{\partial a}\frac{da}{d\alpha} + \frac{\partial \varphi}{\partial b}\frac{db}{d\alpha}. \tag{A}$$

By Leibniz's rule:

$$\frac{\partial \varphi}{\partial \alpha} = \int_{a(\alpha)}^{b(\alpha)} f'_\alpha(x, \alpha)\, dx; \tag{1}$$

by the rule for differentiation of an integral with respect to its upper limit:

$$\frac{\partial \varphi}{\partial b} = f[b(\alpha), \alpha)]; \tag{2}$$

and by the same rule:

$$\frac{\partial \varphi}{\partial a} = -f[a(\alpha), \alpha]. \tag{3}$$

The required equation follows on substituting in (A) for $\partial\varphi/\partial\alpha$ from (1), for $\partial\varphi/\partial b$ from (2), and for $\partial\varphi/\partial\alpha$ from (3).

* We can convince ourselves of this by substituting

$$x = a(\alpha) + [b(\alpha) - a(\alpha)]\, t,$$

and transforming the given integral to one with constant limits 0 and 1.

IV. Integration of an integral with respect to a parameter. We take the integral with respect to α of the function

$$F(\alpha) = \int_{a(\alpha)}^{b(\alpha)} f(x, \alpha)\, dx,$$

continuous in the interval $[\alpha_1, \alpha_2]$:

$$\int_{\alpha_1}^{\alpha_2} F(\alpha)\, d\alpha = \int_{\alpha_1}^{\alpha_2} d\alpha \int_{a(\alpha)}^{b(\alpha)} f(x, \alpha)\, dx.$$

We obtain as a result an iterated integral over a domain in the $Ox\alpha$ plane which depends on the given limits of the integrals. On using the familiar property of iterated integrals (see Sec. 172, Remark 2), we can change the order of integration, i.e. integrate first with respect to the parameter, then with respect to the variable of integration.

In the particular case when limits a and b are constant (i.e. are independent of parameter α), we get

$$\int_{\alpha_1}^{\alpha_2} F(\alpha)\, d\alpha = \int_{\alpha_1}^{\alpha_2} d\alpha \int_{a}^{b} f(x, \alpha)\, dx = \int_{a}^{b} dx \int_{\alpha_1}^{\alpha_2} f(x, \alpha)\, d\alpha.$$

This formula leads to a rule analogous to Leibniz's rule. Both rules can be covered by the same general statement.

LEIBNIZ'S RULE. **To differentiate or integrate an integral with constant limits with respect to a parameter we apply the operation to the integrand.**

Some examples of the use of integration with respect to a parameter may be mentioned.

(1) *Cauchy's formula.* Given the continuous function $y = f(x)$, we successively integrate $f(x)$ n times with variable upper limits and constant lower limit a:

$$\underbrace{\int_a^x dx \int_a^x dx \ \ldots \int_a^x}_{n \text{ times}} f(x)\, dx.$$

This n-ple integral is given very simply by a single integral.

In fact, if $n = 2$ we have by Dirichlet's formula (Sec. 173) (on changing the notation for the variable of integration in the first integral):

$$\int_a^x dx \int_a^x f(t)\, dt = \int_a^x f(t)\, dt \int_t^x dx = \int_a^x (x - t)\, f(t)\, dt.$$

Hence, with $n = 3$:

$$\int_a^x dx \int_a^x dx \int_a^x f(x)\,dx = \int_a^x dx \int_a^x (x-t)\,f(t)\,dt$$

$$= \int_a^x f(t)\,dt \int_t^x (x-t)\,dx$$

$$= \int_a^x f(t)\,\frac{1}{2}(x-t)^2 \Big|_t^x dt = \frac{1}{2}\int_a^x (x-t)^2 f(t)\,dt.$$

We obtain on proceeding in this way:

$$\underbrace{\int_a^a dx \int_a^a dx \cdots \int_a^a}_{n \text{ times}} f(x)\,dx = \frac{1}{(n-1)!}\int_a^a (x-t)^{n-1} f(t)\,dt.$$

This is Cauchy's formula. Incidentally this formula can be checked by differentiating both sides n times with respect to x (we have to use on the right-hand side the rule for differentiation of an integral with respect to a parameter).

(2) *Integration of improper integrals.* The rule for integration of an integral with respect to a parameter also holds for improper integrals if these are uniformly convergent with respect to the parameter. This rule is sometimes used for evaluating improper integrals.

We found in Sec. 181 that

$$\int_0^\infty \frac{\sin \beta x}{x}\,dx = \frac{\pi}{2}, \quad \beta > 0.$$

Since this integral is uniformly convergent in β, integration between limits of variation of β from a to b, $b > a > 0$, gives

$$\int_a^b d\beta \int_0^\infty \frac{\sin \beta x}{x}\,dx = \frac{\pi}{2}(b-a),$$

i.e. in accordance with our rule:

$$\int_0^\infty \frac{dx}{x} \int_a^b \sin \beta x\,d\beta = \frac{\pi}{2}(b-a),$$

whence

$$\int_0^\infty \frac{\cos ax - \cos bx}{x^2}\,dx = \frac{\pi}{2}(b-a).$$

CHAPTER XIII

LINE AND SURFACE INTEGRALS

1. Line Integrals

181. Problems concerning Work. Integrals over an Arc. We have so far considered as domains of integration either a segment of the real axis or domains of the plane or of space.

We shall now take as domain of integration a curve on the plane or in space, or a surface in space*. As a preliminary, we shall consider the problem of work.

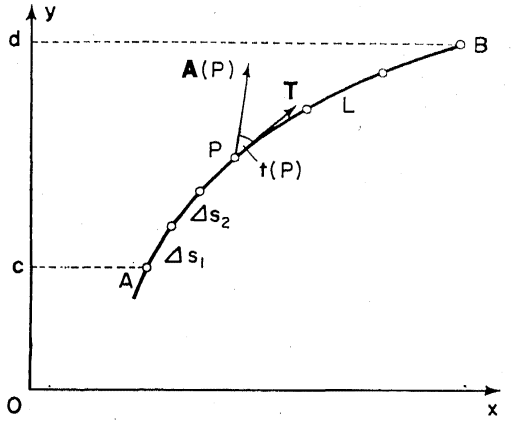

FIG. 51

Let a particle P move along a plane curve $L(AB)$ under the action of a force $\mathbf{A}(P)$ on it, varying in general both in magnitude and direction along the path. Let $|\mathbf{A}(P)| = \varphi(P)$, $(\mathbf{A}(P), \mathbf{T}) = \tau(P)$ (Fig. 51).

Let us find the work done by the force $\mathbf{A}(P)$ over the entire path L.

* We shall not give the existence theorems for the integrals described in this chapter since they can be formulated exactly as in previous cases (see Secs. 86, 167).

We first of all recall the premises from which we started when defining work (see Sec. 85): (1) the work done throughout the path is equal to the sum of the works over the parts into which the path is divided (property of additiveness); (2) the work done by a constant force is measured by the product of the force and the path-length, work being done only by the component of the force in the direction of the motion (the *tangential component*).

We divide curve L into n parts (n pieces of arc) $s_1, s_2, \ldots, s_n$, the lengths of which are denoted by $\Delta s_1, \Delta s_2, \ldots, \Delta s_n$ respectively.

Let us suppose that the force—the vector $\mathbf{A}(P)$—remains constant over a piece of arc s_i, with a value equal to its value at some point P_i of s_i. The tangential component $\mathbf{A}(P)$ is equal to the projection of $\mathbf{A}(P_i)$ on the direction of the tangent P_iT_i, i.e. is equal to $\varphi(P_i)\cos\tau(P_i)$. The work done over the piece of arc s_i is now obviously

$$\varphi(P_i)\cos\tau(P_i)\,\Delta s_i, \quad \text{or} \quad f(P_i)\,\Delta s_i,$$

if we put $\varphi(P_i)\cos\tau(P_i) = f(P_i)$. On adding the amounts of work on all the pieces of arc $s_1, s_2, \ldots, s_n$ we obtain

$$A_n = f(P_1)\,\Delta s_1 + f(P_2)\,\Delta s_2 + \cdots + f(P_n)\,\Delta s_n = \sum_{i=1}^{n} f(P_i)\,\Delta s_i. \quad (*)$$

We take A_n as the approximate value of the required work A, which we define as the limit of A_n as $\max \Delta s_i \to 0$:

$$A = \lim_{\max \Delta s_i \to 0} A_n.$$

Now let any curve L be given in the Oxy plane and let $f(P)$ be a variable function $f(P)$ on L.

The sum (*) is called the n-th *integral sum over the curve for function* $f(P)$ *and curve* L.

Definition. **The limit of integral sum (*) as the piece of arc of greatest length tends to zero is called the *line integral of function* $f(P)$ *over curve* L and is written as**

$$\lim \sum_{i=1}^{n} f(P_i)\,\Delta s_i = \int_L f(P)\,ds.$$

Curve $L(AB)$ is called the *line* or *contour* or *path of integration*; the variable arc s is the *variable of integration*; the function $f(P)$ is the *integrand*, whilst $f(P)\,ds$ is the *element of the integral*. In particular, if the contour L is a segment of the real axis, we arrive

at the concept of *rectilinear* integral over an interval, which precisely corresponds with the general definition of integral given in Sec. 168.

We can therefore say, using the new concept of integral now introduced, that *the work done over a path by a variable force is given by the line integral of the projection of the force on the direction of motion, taken over the path*:

$$A = \int_L \varphi(P) \cos\tau(P)\, ds.$$

An important point is that the path—the curve L—can be of any sort, either *open* or *closed*.

Similar definitions can be given for the work done during motion over a spatial curve and for the line integral in this case.

We observe that *the curves encountered in this chapter are always assumed to be* "piecewise-smooth", *i.e. they consist of a finite number of continuous curves with continuously rotating tangent* (see Sec. 54). This condition will not be stipulated below.

182. Properties, Evaluation and Applications of Line Integrals. The general definition of integral given in Sec. 168 has full reference to the line integral over an arc. The domain of integration W is now the curve L, whilst the measure of sub-domain Δw_i is the length of a piece of arc Δs_i. The properties of the ordinary integral are therefore retained for the integral over an arc. In particular, if the contour of integration—curve L—is divided into n arcs $s_1, s_2, \ldots, s_n$, the integral over L will be equal to the sum of the integrals over the pieces of arc:

$$\int_L f(P)\, ds = \int_{s_1} f(P)\, ds + \int_{s_2} f(P)\, ds + \cdots + \int_{s_n} f(P)\, ds.$$

It must be mentioned that, given the integrand $f(P)$ of a line integral, it is of no consequence which terminal points of the curve are reckoned initial and final points, in other words, in which direction the curve is measured.

We now turn to the method of evaluating a line integral over an arc. This integral will be seen to be reducible to an ordinary (rectilinear) integral.

When the variable point P of curve $L(AB)$ moves from the initial point A to the final point B the length s of arc AP increases monotonically from zero to the total length l of curve AB. For each position of point P there is a corresponding value s, i.e. $f(P)$ can be regarded as a function of s, i.e. $f(P) = F(s)$.

In fact, let

$$x = x(s), \quad y = y(s)$$

be the *natural equations* of curve L (see Sec. 160). On substituting these expressions for x and y in function $f(x, y)$, we get: $f[x(s), y(s)] = F(s)$. Thus

$$I_n = \sum_{i=1}^{n} f(P_i)\,\Delta s_i = \sum_{i=1}^{n} F(s_i)\,\Delta s_i.$$

This is an ordinary integral sum with respect to variable s. On passing to the limit we get an expression for line integral I as an ordinary integral:

$$I = \int_0^l F(s)\,ds. \qquad (**)$$

Curve L is generally given by parametric equations

$$x = x(t), \quad y = y(t),$$

where the parameter t differs from the length of arc s. But s can be reckoned as a function of t, which we shall assume to be continuously differentiable (i.e. to have a continuous derivative) in the interval $[t_1, t_2]$ of variation of t corresponding to the whole of contour L. We now obtain on substituting $s = s(t)$ in integral (**):

$$I = \int_{t_1}^{t_2} \Phi(t)\,\frac{ds}{dt}\,dt,$$

where

$$\Phi(t) = F[s(t)] = f[x(t), y(t)].$$

Since

$$\frac{ds}{dt} = \sqrt{x'^2(t) + y'^2(t)},$$

we can write

$$\int_L f(P)\,ds = \int_{t_1}^{t_2} f[x(t), y(t)]\,\sqrt{x'^2(t) + y'^2(t)}\,dt.$$

RULE FOR EVALUATING A LINE INTEGRAL OVER AN ARC. **To transform the line integral over the arc**

$$x = x(t), \quad y = y(t)$$

into an ordinary integral we have to replace x, y and ds in an element of the integral by their expressions in t and dt, and integrate over the interval of variation of t corresponding to the contour of integration.

If the parametric equations are different for different pieces of the contour of integration L, we must divide it into separate parts and work out the entire integral as the sum of the integrals over the parts.

The general rule for reducing a line integral to an ordinary integral leads to reduction rules for the special cases when $t = x$ or $t = y$. For instance, if the integral over curve AB (Fig. 52) is to be evaluated with the aid of rectilinear integrals with respect to x, curve AB has to be divided into three parts: AC, CD, DB. Each of these curves has an equation of the form $y = y(x)$, where $y(x)$ is a single-valued function, and the integral over AB is expressed as the sum of three ordinary integrals.

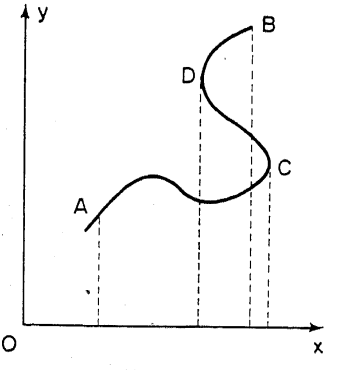

FIG. 52

For the line integral over the spatial curve

$$x = x(t), \quad y = y(t), \quad z = z(t)$$

we have:

$$\int_L f(P)\,ds = \int_{t_1}^{t_2} f[x(t), y(t), z(t)] \times$$

$$\times \sqrt{x'^2(t) + y'^2(t) + z'^2(t)}\,dt.$$

The evaluation rule here is similar to that for a plane curve.

The importance of the concept of line integral is obvious from the following: in order to find some quantity corresponding to a curve L (say the work done by a force along the path L) it may be necessary to take *several* ordinary integrals corresponding to separate sections of the curve, whereas this can be done with the aid of a *single* line integral over the entire contour. An important point here is that the line integral has all the basic properties of the ordinary integral.

The present type of integral is applied to concrete problems precisely as in other cases, by finding the differential of the required magnitude and "summing" it over the length of arc.

Apart from the problem of work considered in Secs. 148 and 181, we shall take a geometrical problem, the solution of which becomes particularly simple with the aid of the line integral. Let G be a cylindrical surface (Fig. 53), the cross-section of which is the curve L in Oxy and the generators of which are perpendicular to this plane. We shall find the area Q of the part of the surface included

between curve L and any curve L_1 lying above L on the same surface.

The "height" of the surface corresponding to point P of base L (i.e. the z co-ordinate of point M of curve L_1) is a function of the point P of L:

$$z = f(P) = f(x, y).$$

We take an infinitesimal element of arc ds at an arbitrary point P of the base. The surface area Δq corresponds to this, the principal

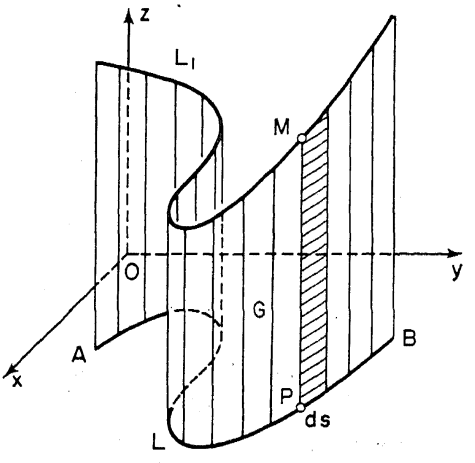

FIG. 53

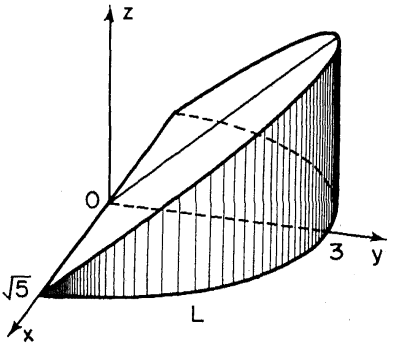

FIG. 54

part of Δq, proportional to ds, being the area of the rectangle with base ds and height equal to the height of the surface at point P. The differential dq of the required quantity is therefore

$$dq = f(P)\,ds.$$

We obtain on summing these "elementary" areas over curve L:

$$Q = \int_L f(P)\,ds.$$

Example. Find the area Q of the lateral surface of half the elliptic cylinder $x^2/5 + y^2/9 = 1$, $y \geqslant 0$, $z \geqslant 0$, cut by the plane $z = y$ (Fig. 54).

We have:

$$Q = \int_L z\,ds = \int_L y\,ds,$$

where L is the arc of the ellipse $x^2/5 + y^2/9 = 1$, $y \geqslant 0$.

Let us evaluate the integral. The parametric equations of the contour of integration:

$$x = \sqrt{5}\cos t, \quad y = 3\sin t$$

give

$$Q = \int_0^{\pi} 3\sin t \sqrt{9\cos^2 t + 5\sin^2 t}\, dt = 3\int_0^{\pi} \sin t \sqrt{4\cos^2 t + 5}\, dt.$$

We find on setting $\cos t = u$:

$$Q = -3\int_1^{-1} \sqrt{4u^2 + 5}\, du = 6\int_0^1 \sqrt{4u^2 + 5}\, du$$

$$= 12\left[\frac{u}{2}\sqrt{u^2 + \frac{5}{4}} + \frac{5}{8}\ln\left(u + \sqrt{u^2 + \frac{5}{4}}\right)\right]\Bigg|_0^1 = 9 + \frac{15}{4}\ln 5.$$

2. Co-ordinate Line Integrals

183. Co-ordinate Line Integrals.

I. Definition and properties. We take a curve L in the Oxy plane and a continuous function $f(P) = f(x, y)$, specified in some domain that includes the curve L. We divide L into n arbitrary pieces and let P_i be any point of the i-th arc. We now multiply $f(P_i)$, not by the element of arc Δs_i of the curve, but by the element Δx_i of Ox corresponding to the element of arc (Fig. 55): $f(P_i)\,\Delta x_i$. Finally, we form the sum I_n of all these products:

$$I_n = \sum_{i=1}^{n} f(P_i)\,\Delta x_i; \qquad (*)$$

the limit of I_n as $\max|\Delta x_i| \to 0$:

$$I = \lim_{\max|\Delta x_i|\to 0} \sum_{i=1}^{n} f(P_i)\,\Delta x_i$$

is written as

$$I = \int_L f(P)\,dx = \int_L f(x, y)\,dx. \qquad (**)$$

The sum (*) is called *the n-th integral sum with respect to the x co-ordinate for function $f(P)$ and curve L.*

Definition. **The limit I (**) of integral sum I_n (*) as the maximum sub-interval of Ox tends to zero is termed *the line integral with***

respect to the x co-ordinate* (or simply with respect to *x*) *of function f(P) over curve L.

The line integral with respect to y is similarly defined:

$$\lim_{\max|\Delta y_i|\to 0}\sum_{i=1}^{n} f(P_i)\,\Delta y_i = \int_L f(P)\,dy = \int_L f(x, y)\,dy.$$

The same terminology as above is used for these integrals, though the variable of integration is now x or y instead of arc s.

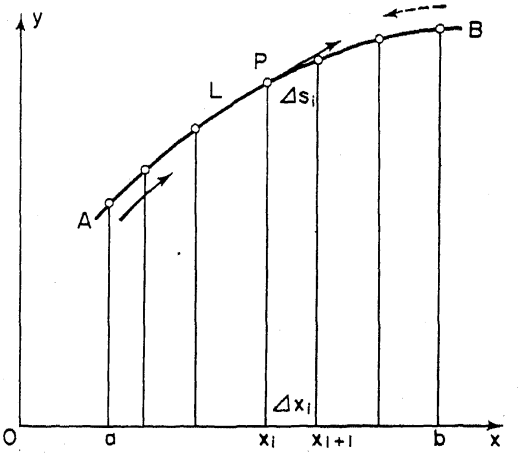

FIG. 55

Whereas the sense of the circuit along L is of no importance in the line integral over the arc, it plays an essential role as regards the co-ordinate line integral.

Two opposite directions (orientations) can be chosen on any curve; thus curve L (Fig. 55) can be orientated from point A to point B or vice versa. When specifying a co-ordinate line integral we have to indicate not only the domain of integration—the curve L—but also its orientation, i.e. the direction in which the integration is performed.

If the direction of integration is changed the co-ordinate line integral takes the opposite sign.

For, when the sense of the circuit along L changes, all the elements Δx (Δy) in the integral sum change sign, so that we have, as in the case of the corresponding property of the ordinary integral (Sec. 89):

$$\int_{+L} f(P)\,dx = -\int_{-L} f(P)\,dx \quad \text{and} \quad \int_{+L} f(P)\,dy = -\int_{-L} f(P)\,dy,$$

where $+L$ and $-L$ denote curve L in its two opposite senses.

An anti-clockwise circuit round a plane closed curve is described as positive. The domain bounded by the curve remains on the left when the circuit is in the positive sense. It must be noted that (with the usual disposition of the co-ordinate axes) this sense is the same as that in which the positive semi-axis Ox rotates via the shortest path to becomes positive Oy.

The ordinary (rectilinear) integral is a particular case of a co-ordinate line integral when the contour of integration is part of a co-ordinate axis.

Co-ordinate line integrals are closely connected with integrals over an arc. In fact, since

$$\frac{dx}{ds} = \cos\alpha(P), \quad \frac{dy}{ds} = \sin\alpha(P),$$

where $\alpha(P)$ is the angle between the orientated curve L at point P and Ox (see Sec. 163), we have

$$\int_L f(P)\,dx = \int_L f(P)\cos\alpha(P)\,ds$$

and

$$\int_L f(P)\,dy = \int_L f(P)\sin\alpha(P)\,ds = \int_L f(P)\cos\beta(P)\,ds,$$

where $\beta(P) = \frac{1}{2}\pi - \alpha(P)$ is the angle between L at the point P and Oy.

Conversely:

$$\int_L f(P)\,ds = \int_L f(P)\,\frac{1}{\cos\alpha(P)}\,dx,$$

$$\int_L f(P)\,ds = \int_L f(P)\,\frac{1}{\cos\beta(P)}\,dy.$$

Co-ordinate line integrals possess all the basic properties of the ordinary integral.

We shall dwell on an important specific property.

THEOREM. **If a domain D bounded by a closed curve L is divided into two parts D_1 and D_2, the line integral over the whole of L can be written as the sum of the integrals in the same sense** (say positive) **over curves L_1 and L_2 bounding domains D_1 and D_2** (we assume that the integrand is defined and continuous throughout D).

Proof. Let domain D be bounded by the closed curve L ($AECBA$), whilst domains D_1 and D_2 are bounded respectively by L_1 ($AECA$) and L_2 ($ACBA$) (Fig. 56). We have:

$$\int_{L_1} f(P)dx = \int_{AEC} f(P)dx + \int_{CA} f(P)dx$$

and

$$\int_{L_2} f(P)dx = \int_{AC} f(P)dx + \int_{CBA} f(P)dx;$$

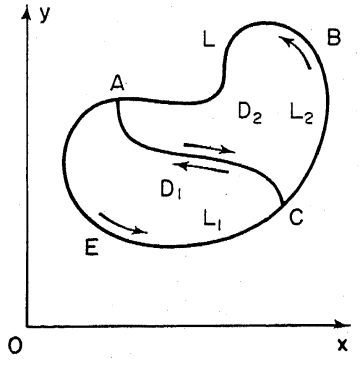

Fig. 56

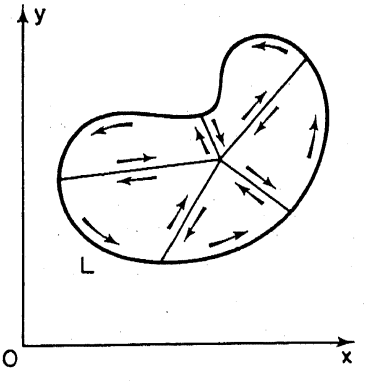

Fig. 57

the integrals over CA and AC are taken over the same curve but in opposite directions; their sum therefore vanishes. On adding the two equations term by term, we get

$$\int_{L_1} f(P)dx + \int_{L_2} f(P)dx = \int_{AEC} f(P)dx + \int_{CBA} f(P)dx = \int_{L} f(P)dx.$$

This is what we had to prove.

As the reader may easily verify, this theorem holds for any number of domains into which domain D may be divided:

$$\int_{L} f(P)dx = \int_{L_1} f(P)dx + \int_{L_2} f(P)dx + \cdots + \int_{L_n} f(P)dx,$$

all the closed contours of integration L and L_1, L_2, ..., L_n having the same sense with respect to the domains bounded by them (Fig. 57).

II. Evaluation of the line integral. As in the previous case, the co-ordinate line integral is usually evaluated by transforming it to an ordinary integral.

We state the general rule without a repetition of the proof.

RULE FOR EVALUATING A CO-ORDINATE LINE INTEGRAL. **To transform such an integral over the curve $x = x(t)$, $y = y(t)$ into an ordinary integral, we must replace x, y and dx (or dy) in the integrand element by their expressions in terms of t and dt and take the integral over the interval $[t_1, t_2]$ of the t axis corresponding to the contour of integration:**

$$\int_{(A)}^{(B)} f(x, y)\, dx = \int_{t_1}^{t_2} f[x(t), y(t)]\, x'(t)\, dt,$$

$$\int_{(A)}^{(B)} f(x, y)\, dy = \int_{t_1}^{t_2} f[x(t), y(t)]\, y'(t)\, dt,$$

where t_1 and t_2 are the values of t corresponding to points A and B of curve L.

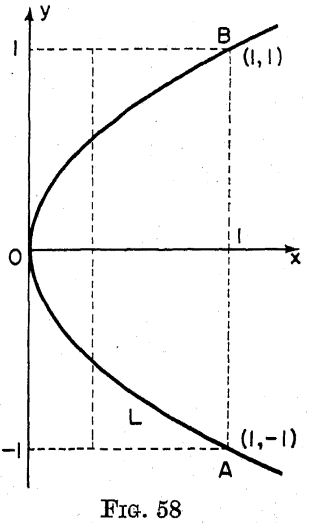

FIG. 58

In particular, this leads us to rules for the cases $t = x$ and $t = y$. If the equation of the complete contour cannot be given with the aid of single-valued functions, the contour must be sub-divided and the integrals evaluated over the separate parts.

Example. Let us evaluate

$$I = \int_L xy\, dx,$$

where L is the arc of the parabola $y^2 = x$ from the point $(1, -1)$ to $(1, 1)$ (Fig. 58).

Since the function $y(x)$ is not single-valued in the equation of the complete contour when solved for y: $y = \pm\sqrt{x}$, we split L into two parts: AO and OB, the equations of which are $y = -\sqrt{x}$ and $y = +\sqrt{x}$. Now,

$$I = \int_{(A)}^{(0)} xy\, dx + \int_{(0)}^{(B)} xy\, dx$$

$$= -\int_1^0 x^{\frac{1}{2}} x\, dx + \int_0^1 x^{\frac{1}{2}} x\, dx = 2\int_0^1 x^{\frac{3}{2}}\, dx = \frac{4}{5}.$$

The integral of xy over the same contour but with respect to the y co-ordinate is zero:

$$\int_L xy\, dy = \int_{-1}^{1} y^2 y\, dy = 0.$$

Similar definitions and properties can be given for co-ordinate line integrals in the case of a spatial curve. Three integrals can be taken of a given function $f(P) = f(x, y, z)$ over a given curve $L(AB)$:

$$\int_L f(P)\,dx, \quad \int_L f(P)\,dy, \quad \int_L f(P)\,dz.$$

Co-ordinate line integrals over a spatial curve are closely connected with integrals over an arc. Let $\alpha = \alpha(P), \beta = \beta(P), \gamma = \gamma(P)$ be the angles of orientation of curve L at point P with axes Ox, Oy, Oz; then

$$\int_L f(P)\,dx = \int_L f(P)\cos\alpha(P)\,ds,$$

$$\int_L f(P)\,dy = \int_L f(P)\cos\beta(P)\,ds,$$

$$\int_L f(P)\,dz = \int_L f(P)\cos\gamma(P)\,ds.$$

Let $x = x(t)$, $y = y(t)$, $z = z(t)$ be the parametric equations of curve L. Line integrals over L are transformed to ordinary integrals by the formulae

$$\int_L f(x, y, z)\,dx = \int_{t_1}^{t_2} f[x(t), y(t), z(t)]\,x'(t)\,dt \quad \text{etc.}$$

where t_1 and t_2 are the values of t corresponding to the initial and final points of L. The evaluation rule is similar to that for a plane curve.

REMARK. If the integrand $f(P)$ is defined, or looked on as defined, only on the curve L, the line integral (with respect to arc or co-ordinate) along L is in essence merely a convenient, unified method of writing one or more ordinary integrals. It is quite another matter when $f(P)$ is defined and continuous in a domain containing various curves as paths of integration; the line integral (with respect to arc or co-ordinate)

$$\int_L f(P)\,ds \quad \text{or} \quad \int_L f(P)\,dx \quad \text{etc.}$$

takes values dependent on the contour of integration. For each such contour there is a corresponding value of the integral; the line integral is said to be a *functional in the domain.*

In general *a magnitude is termed a* functional *if there is a definite value of it corresponding to each curve in question* (i.e. to each function or system of functions).

It is of considerable interest to investigate the behaviour of a line integral as a functional, i.e. to see how it varies as the path of integration varies. The elements of this study will be given below.

184. Component Line Integrals. Green's Formula.

I. In mathematical physics and its applications great importance attaches to combinations (sums) of line integrals such as

$$\int_L X\,dx + \int_L Y\,dy,$$

where $X = X(x, y)$, $Y = Y(x, y)$. They are written with the single integral sign:

$$\int_L X\,dx + Y\,dy.$$

Similar sums are often encountered in space:

$$\int_L X\,dx + Y\,dy + Z\,dz,$$

where $X = X(x, y, z)$, $Y = Y(x, y, z)$, $Z = Z(x, y, z)$.

All the properties and evaluation methods described in Sec. 183 apply to these line integrals.

We take as a first example the problem of work discussed in Sec. 181.

Element $\varphi(P)\cos\tau(P)\,ds$ of the integral expressing the work is in fact *the scalar product of the force vector* $\mathbf{A}(P)$ *at point* P *with the infinitesimal displacement vector* $\overline{dP}$.

We therefore write conditionally:

$$A = \int_L \boldsymbol{A}(P) \cdot \overline{dP}.$$

But instead of specifying the force at point P, i.e. vector $\mathbf{A}(P)$, by its magnitude $\varphi(P)$ and inclination $\tau(P)$ to the direction of motion, we can specify it by its projections $X = X(x, y)$, $Y = Y(x, y)$ on Ox and Oy. Since the projections of vector $\overline{dP}$ on Ox and Oy are dx and dy, the familiar expression for the scalar product gives us

$$dA = \boldsymbol{A}(P) \cdot \overline{dP} = X(x, y)\,dx + Y(x, y)\,dy.$$

Consequently,

$$A = \int_L X(x, y)\,dx + Y(x, y)\,dy,$$

the integration being in the direction of motion.

Similarly, we have for the work done over a spatial curve:

$$A = \int_L X(x, y, z)\,dx + Y(x, y, z)\,dy + Z(x, y, z)\,dz,$$

where X, Y, Z are the projections of the force at point P of curve L on Ox, Oy, Oz.

II. We shall prove an important theorem of analysis.

GREEN'S THEOREM*. **If functions $X(x, y)$, $Y(x, y)$ are continuous along with their first partial derivatives in domain D, we have**

$$\iint_D \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right) dx\, dy = \int_L X\, dx + Y\, dy, \qquad (*)$$

where L is the boundary of D and the integration is in the positive direction along L.

Expression (*) is called *Green's formula*.

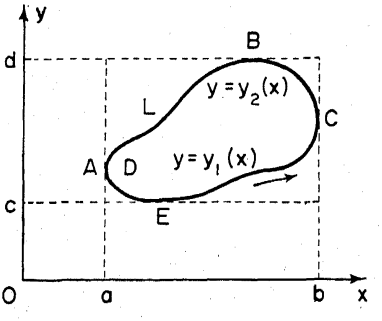

FIG. 59

Proof. We first take a singly-connected (see Sec. 139) domain D in the Oxy plane, bounded by a curve L which is cut by lines parallel to the axes in not more than two points (Fig. 59). We shall transform the double integral

$$I = \iint_D \frac{\partial X}{\partial y}\, dx\, dy.$$

Integration with respect to y, then with respect to x, gives

$$I = \int_a^b dx \int_{y_1(x)}^{y_2(x)} \frac{\partial X}{\partial y}\, dy,$$

where $y = y_2(x)$ is the equation of curve ABC, $y = y_1(x)$ is the

* G. GREEN (1793–1841), a well known English physicist and mathematician. Green's formula is a particular case of the general formula obtained by M. V. OSTROGRADSKII (see Introduction, Sec. 4).

equation of curve AEC. We obviously have, on carrying out the inner integration:

$$I = \int_a^b \{X[x, y_2(x)] - X[x, y_1(x)]\}\, dx$$

$$= \int_a^b X[x, y_2(x)]\, dx + \int_b^a X[x, y_1(x)]\, dx.$$

But the right-hand side is the line integral with respect to x of function $X(x, y)$ over contour L in the direction $ABCE$.

Thus

$$\iint_D \frac{\partial X}{\partial y}\, dx\, dy = \int_{-L} X\, dx,$$

the circuit round L being in the negative direction (clockwise).

We can prove in the same way that

$$\iint_D \frac{\partial Y}{\partial x}\, dx\, dy = -\int_{-L} Y\, dy.$$

On subtracting the latter equation from the former and changing the direction of integration, we obtain Green's formula:

$$\iint_D \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right) dx\, dy = \int_{+L} X\, dx + Y\, dy,$$

where the line integral is now taken in the positive direction round L.

Our assumption regarding L does not exclude the possibility of it containing segments of straight lines parallel to the axes. The reader will observe that there is no change in the proof here, since the parts of line integrals $\int_L X\, dx$, $\int_L Y\, dy$ relating to straight line segments $x = \text{const.}$, $y = \text{const.}$ respectively are zero, due to the fact that $dx = 0$ or $dy = 0$ on them. In particular, formula (*) holds for curvilinear trapezia.

We show now that (*) holds for domains bounded by *any* closed curve L. In fact, if the boundary does not satisfy the condition originally stipulated (that it is cut by co-ordinate lines in not more than two points), domain D can be sub-divided into $D_1, D_2, \ldots, D_n$, such that their boundaries $L_1, L_2, \ldots, L_n$ satisfy the condition (Fig. 60).

If we write $I_1, I_2, \ldots, I_n$ for the double integrals of the function $\partial Y/\partial x - \partial X/\partial y$ over the domains $D_1, D_2, \ldots, D_n$ respectively, we have

$$I = I_1 + I_2 + \cdots + I_n$$

$$= \int_{L_1} X\,dx + Y\,dy + \int_{L_2} X\,dx + Y\,dy + \cdots + \int_{L_n} X\,dx + Y\,dy.$$

But, by the property of line integrals (Sec. 183), the sum of the integrals on the right-hand side is equal to the integral over the whole of boundary L. This is what we had to show.

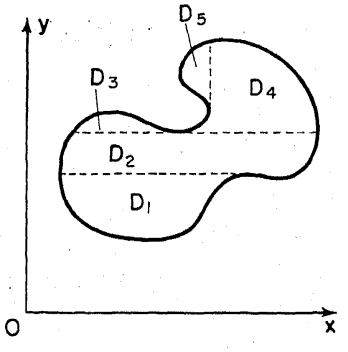

FIG. 60

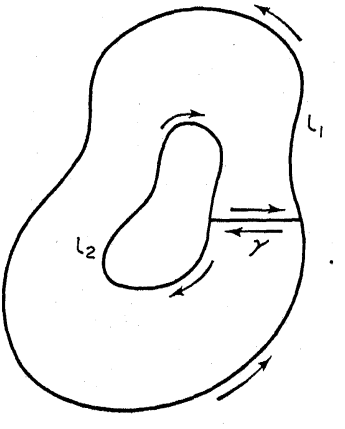

FIG. 61

Now let domain D be *multiply-connected*. Green's formula (*) remains valid in this case; though we have to remember that the line integral on the right-hand side is taken over the entire boundary in such a sense that the domain remains on the left. For instance, suppose we take a doubly-connected domain (Fig. 61). We join outer boundary l_1 and inner boundary l_2 by means of a straight "cut" γ. This gives a singly-connected domain D' bounded by the contour $L' = l_1 + l_2 + \gamma$. Since formula (*) has been proved for this, we can write

$$\iint_{D'} \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right) dx\,dy = \int_{L'} X\,dx + Y\,dy,$$

where domain D' remains on the left during the circuit round L' (indicated by arrows in Fig. 61). In view of the fact that the cut is traversed in opposite directions and that the double inte-

gral over D' is equal to the same integral over D (see Sec. 179), we get

$$\iint_D \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right) dx\, dy = \int_{l_1} X\, dx = Y\, dy + \int_{-l_2} X\, dx + Y\, dy$$
$$= \int_L X\, dx + Y\, dy,$$

where L denotes the entire contour bounding D. This contour consists of an inner and outer contour ($L = l_1 + l_2$ in Fig. 61); the circuit round the outer contour is positive during integration and negative round the inner.

Green's theorem is now fully proved.

Green's formula enables a double integral, i.e. an integral over a plane domain, to be replaced by a line integral over its boundary, and conversely, a line integral round a closed curve can be replaced by a double integral over the domain bounded by the curve.

In particular, let $X = -y$, $Y = x$. Then $\partial Y/\partial x = 1$, $\partial X/\partial y = -1$, i.e.

$$S = \iint_D dx\, dy = \frac{1}{2}\int_L - y\, dx + x\, dy,$$

i.e. we have obtained an *expression for the area of domain D in terms of a line integral.*

Example. Find the area of a loop of the *folium of Descartes:*

$$x^3 + y^3 - 3axy = 0$$

(see Sec. 75). We can take as parametric equations of the folium:

$$x = \frac{3at}{1 + t^3}, \quad y = \frac{3at^2}{1 + t^3} = xt.$$

Parameter $t\ (= y/x)$ is equal to the tangent of the angle between Ox and the radius vector of the point of the curve: $t = \tan\alpha$. Point $P(x, y)$ thus traces out the entire loop in a positive sense (anticlockwise) when t varies monotonically from $t = 0$ to $t = \infty$.

We transform the line integral for the required area to the ordinary integral. The working is simplified if we leave the differentiations for obtaining dx and dy to the end:

$$S = \frac{1}{2}\int_L x\, dy - y\, dx = \frac{1}{2}\int_0^\infty [x(x + tx') - xtx']\, dt = \frac{1}{2}\int_0^\infty x^2 dt,$$

i.e.

$$S = \frac{9a^2}{2}\int_0^\infty \frac{t^2\,dt}{(1+t^3)^2} = \frac{3a^2}{2}\int_0^\infty \frac{d(t^3)}{(1+t^3)^2} = \frac{3a^2}{2}\left(-\frac{1}{1+t^3}\right)\Bigg|_0^\infty = \frac{3a^2}{2}.$$

185. Independence of the Integral on the Contour of Integration. Let $X(x, y)$, $Y(x, y)$ be *continuous functions* given in a *singly-connected* domain D of the Oxy plane. We take the line integral

$$I_{P_0P} = \int_{(P_0)}^{(P)} X(x, y)\,dx + Y(x, y)\,dy \qquad (*)$$

over a curve L in domain D joining two given points $P_0(x_0, y_0)$ and $P(x, y)$ of domain D.

The value of integral I_{P_0P} in general depends on the *contour of integration* L. For instance, integral

$$\int_{(0,0)}^{(1,1)} x\,dy - y\,dx \quad (X = -y, Y = x),$$

over the parabola $y = x^2$ is equal to $\frac{1}{3}$, and over the cubical parabola $y = x^3$ is equal to $\frac{1}{2}$.

The question arises as to the *conditions to be imposed on functions* $X(x, y)$ *and* $Y(x, y)$ *so that integral* (*) *does not depend on the path of the integration but only on the initial and final points* $P_0(x_0, y_0)$ *and* $P(x, y)$ *of the path.*

The importance of this problem will be clear from the following. Functions X and Y might be say the projections of a force; integral (*) would then give the work done in a displacement from point P_0 to P. The independence of the integral on the part of integration thus implies that the work is the same no matter what the path of the displacement from point P_0 to P under the action of a force, this latter being evidently an important characteristic of the force.

The present problem is solved by the following fundamental theorem.

THEOREM. **The necessary and sufficient condition for independence of line integral (*) of the contour contained in a singly-connected domain D and joining given points P_0 and P is that functions $X(x, y)$, having continuous partial derivatives in domain D, satisfy**

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x} \qquad (A)$$

at every point of D.

We observe first of all that *asserting the independence of the integral on the path of integration in a domain D is equivalent to asserting the vanishing of the integral over any closed path in D**. For suppose that integral (*) is known to be independent of the path of integration; we can now show that it vanishes over any closed path. Let L be a closed contour, and P_0 and P two points on L (Fig. 62). Since the integral over curve P_0MP is equal by hypothesis to the integral over P_0NP:

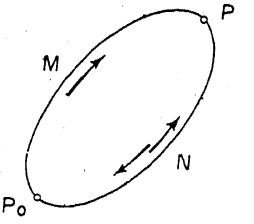

FIG. 62

$$I_{P_0MP} = I_{P_0NP}, \quad \text{i.e.} \quad I_{P_0MP} - I_{P_0NP} = 0,$$

we have

$$I_{P_0MP} + I_{PNP_0} = 0,$$

so that

$$I_{P_0MPNP_0} = 0.$$

Conversely, suppose we know that integral (*) vanishes over any closed contour; we show that it is independent of the path of integration. We take any two curves P_0MP and P_0NP (Fig. 62) joining two given points P_0 and P. Since the integral over the closed curve P_0MPNP_0 vanishes by hypothesis:

$$I_{P_0MPNP_0} = 0, \quad \text{i.e.} \quad I_{P_0MP} + I_{PNP_0} = 0,$$

we have

$$I_{P_0MP} = -I_{PNP_0} = I_{P_0NP}.$$

Our fundamental theorem can therefore be stated as:

If functions $X(x, y)$ and $Y(x, y)$ have continuous partial derivatives in a singly-connected domain D, the fact that

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x} \tag{A}$$

at every point of D is necessary and sufficient for integral (*) to vanish over any closed path in D.

Proof. Let l be a closed contour in domain D. In view of the hypothesis, we can apply Green's formula:

$$\int_l X(x, y)\,dx + Y(x, y)\,dy = \iint_\delta \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right) dx\,dy, \quad (**)$$

where δ is the domain bounded by contour l. The sufficiency of condition (A) follows at once from this: if it holds, the double integral, and hence also the line integral in formula (**) vanishes.

* This proposition holds for integrals over both plane and spatial curves.

The necessity of condition (A) is proved by "reductio ad absurdum". Let integral (*) vanish over any closed path in domain D whilst condition (A) *is not fulfilled* at a point P of D, i.e. $\partial Y/\partial x - \partial X/\partial y = \mu \neq 0$. Let say $\mu > 0$. In view of the continuity of the partial derivatives, there must be a δ-neighbourhood of point P such that in it $\partial Y/\partial x - \partial X/\partial y \geqslant \mu - \varepsilon > 0$, where ε is a previously assigned positive number. Using Green's formula (**) and the theorem on the bounds of a double integral (Sec. 169), we have

$$\int_l X(x, y)\,dx + Y(x, y)\,dy = \iint_\delta \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right) dx\,dy \geqslant (\mu - \varepsilon) \cdot \text{area } \delta > 0,$$

where l is the boundary of domain δ. But this contradicts our assumption that the integral vanishes for any closed contour, i.e. $\partial Y/\partial x - \partial X/\partial y$ must vanish identically in domain D.

This completes the proof.

It must be noted that singly-connectedness of domain D is an essential condition for the validity of the theorem. Use has obviously been made in our proof of the singly-connectedness of D; if D is not singly-connected, condition (A) is no longer sufficient for the vanishing of integral (*) over any closed contour.

Let us take an example: $X = -y/(x^2 + y^2)$, $Y = x/(x^2 + y^2)$. These functions are continuous along with their partial derivatives in any circular domain with centre at the origin and excluding the origin. Relationship (A) now holds at every point different from (0, 0):

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x} = \frac{y^2 - x^2}{(x^2 + y^2)^2}.$$

All the conditions of the theorem are thus fulfilled, but in a doubly, not a singly-connected domain D, given by the inequalities $0 < x^2 + y^2 \leqslant R$, where R is the radius of the circle. It may easily be seen that there are closed paths L belonging to domain D over which integral (*) does not vanish. The concentric circles $x^2 + y^2 = r^2$ are examples of such paths. Let $r = 1$; we now have, on putting $x = \cos\varphi$, $y = \sin\varphi$:

$$\int_{x^2+y^2=1} -\frac{y}{x^2 + y^2}\,dx + \frac{x}{x^2 + y^2}\,dy = \int_0^{2\pi} (\sin^2\varphi + \cos^2\varphi)\,d\varphi = 2\pi.$$

A simple explanation of this result will be given later (Sec. 186).

Thus our basic theorem holds only for a singly-connected domain, i.e. a domain bounded by a single closed contour inside which there are no points at which all the hypotheses are not fulfilled, e.g. continuity of functions X and Y; we say that there must be no "holes" in the domain. If there are holes, integration over a path encircling the holes can lead to a non-vanishing value of the integral even though condition (A) is satisfied.

A similar fundamental theorem on independence of the path of integration holds for integrals over spatial curves.

THEOREM. **The necessary and sufficient condition for independence of the integral**

$$\int_{(P_0)}^{(P)} X(x, y, z)\,dx + Y(x, y, z)\,dy + Z(x, y, z)\,dz$$

of the contour of integration belonging to a singly-connected domain Ω and joining given points P_0 and P (or what amounts to the same thing, for its vanishing over any closed path) is that functions $X(x, y, z)$, $Y(x, y, z)$ and $Z(x, y, z)$, continuous along with their partial derivatives in domain Ω, satisfy at every point of the domain:

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x}, \quad \frac{\partial X}{\partial z} = \frac{\partial Z}{\partial x}, \quad \frac{\partial Y}{\partial z} = \frac{\partial Z}{\partial y}. \tag{A_1}$$

We understand by "singly-connected domain in space" a part of space which is bounded by a single closed surface and such that any closed curve in the domain can be contracted to a point.

We leave the proof of the theorem till Sec. 190, where it appears as a simple consequence of Stokes' formula.

186. The Total Differential Test. Alternative Statements of the Fundamental Theorem.

I. Condition (A) of Sec. 185:

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x} \tag{A}$$

implies that

$$X(x, y)\,dx + Y(x, y)\,dy \tag{*}$$

is the *total differential* of a function $I(x, y)$ of two independent variables. We thus have the following theorem.

THEOREM. **The necessary and sufficient condition for differential expression (*), where functions X and Y have continuous partial de-**

rivatives in domain D, to be the total differential of a function $I(x, y)$ in domain D is that equation (A) be satisfied at every point of D.

Proof. The *necessity* follows directly from the fact that, if

$$X(x, y)\,dx + Y(x, y)\,dy = dI(x, y),$$

which is equivalent to the equations

$$X = \frac{\partial I}{\partial x} \quad \text{and} \quad Y = \frac{\partial I}{\partial y},$$

we have by the theorem on the equality of the second mixed derivatives (Sec. 153):

$$\frac{\partial X}{\partial y} = \frac{\partial^2 I}{\partial x\,\partial y} = \frac{\partial^2 I}{\partial y\,\partial x} = \frac{\partial Y}{\partial x}.$$

We now prove the *sufficiency* of condition (A). Suppose that (A) holds in a δ-neighbourhood of point P lying wholly in domain D. By the theorem of Sec. 185, the integral

$$\int_{(P_0)}^{(P)} X(x, y)\,dx + Y(x, y)\,dy,$$

where $P_0(x_0, y_0)$ and $P(x, y)$ are points of domain δ, is independent of the path in this domain that joins points P_0 and P. The integral thus only depends—for a fixed point P_0—on point P, i.e. is a function of P. We write this as $I(P)$:

$$I(P) = \int_{(P_0)}^{(P)} X(x, y)\,dx + Y(x, y)\,dy, \qquad (**)$$

and show that $I(x, y)$ is in fact the function whose total differential in domain δ is equal to expression (*). All we need do is find the partial derivatives of $I(x, y)$ with respect to x and y and verify that they are in fact equal to $X(x, y)$ and $Y(x, y)$ respectively.

We move the final point of the contour of integration by an amount h in the direction of Ox. It takes up the new position $P_1(x + h, y)$. Let h be so small that the whole of the rectilinear segment PP_1 lies in domain δ. Now,

$$I(P_1) = \int_{(P_0)}^{(P_1)} X\,dx + Y\,dy$$

and

$$I(P_1) - I(P) = \int_{(P_0)}^{(P_1)} X\,dx + Y\,dy - \int_{(P_0)}^{(P)} X\,dx + Y\,dy$$
$$= \int_{(P)}^{(P_1)} X\,dx + Y\,dy.$$

The latter integral can be taken over *any* curve joining points $P(x, y)$ and $P_1(x + h, y)$, and in particular, over the *straight line* PP_1. We have $y = \text{const}$ and $dy = 0$ along PP_1. Hence

$$I(P_1) - I(P) = I(x + h, y) - I(x, y) = \int_x^{x+h} X(x, y)\,dx.$$

We now proceed as when seeking the derivative of an ordinary integral with respect to its upper limit. In fact:

$$\frac{I(x + h, y) - I(x, y)}{h} = \frac{1}{h}\int_x^{x+h} X(x, y)\,dx = X(x + \theta h, y),$$

$0 < \theta < 1$ (mean value theorem); on letting h tend to zero, we get

$$\frac{\partial I}{\partial x} = X(x, y).$$

It may be shown similarly that

$$\frac{\partial I}{\partial y} = Y(x, y).$$

This is what we had to prove.

Thus the differential expression in two independent variables $X(x, y)\,dx + Y(x, y)\,dy$ is the differential of a function $I(x, y)$ when and only when* condition (A) is fulfilled. The function $I(x, y)$ is called the primitive of expression $X\,dx + Y\,dy$ and can be found to an accuracy of an arbitrary constant from (**):

$$I(x, y) = \int_{(x_0, y_0)}^{(x, y)} X(x, y)\,dx + Y(x, y)\,dy. \qquad (**)$$

* As opposed to the case of two independent variables, a differential expression of one independent variable (i.e. an expression of the form $X(x)\,(d\,x)$ is always the differential of some function $I(x)$. This function may be called the primitive of $X(x)$ and is found, apart from an arbitrary constant, from the equation

$$I(x) = \int_{x_0}^{x} X(x)\,dx.$$

If condition (A) is not fulfilled, the expression $X\,dx + Y\,dy$ has no primitive.

II. A different statement of the last theorem, which is more convenient in many respects, may be obtained by combining it with the fundamental theorem of Sec. 185:

THEOREM. **Let functions $X(x, y)$ and $Y(x, y)$ be continuous along with their partial derivatives* in a singly-connected domain D; then the necessary and sufficient condition for independence of the line integral**

$$\int_{P_0}^{P} X(x, y)\,dx + Y(x, y)\,dy$$

on the contour of integration in D (or what amounts to the same thing, for its vanishing over any closed path) is that the integrand element $X\,dx + Y\,dy$ be the total differential of some function $I(x, y)$ in domain D:

$$X\,dx + Y\,dy = dI.$$

If all the conditions of the theorem are satisfied except at certain points ("holes") belonging to the singly-connected domain, the theorem may prove to be invalid: the integral over closed paths encircling the "holes" may be non-zero.

This is explained by the fact that the primitive function

$$I(x, y) = \int_{(x_0, y_0)}^{(x, y)} X\,dx + Y\,dy,$$

whilst being *single-valued* in a singly-connected domain, may be *many-valued* in a multiply-connected domain, i.e. its value may be determined not only by the point (x, y) but also by the path along which this point is approached from the initial point (x_0, y_0).

For example, the expression (see Sec. 185)

$$-\frac{y}{x^2 + y^2}\,dx + \frac{x}{x^2 + y^2}\,dy$$

is the total differential of the function $I = \arctan y/x$ at every point of the doubly-connected domain $0 < x^2 + y^2 \leqslant 1$ (the unit circle less its centre), no matter what branch of arc tan y/x we take as I. Yet it was shown in Sec. 185 that the line integral of this differential expression over the circle $x^2 + y^2 = 1$ is equal to 2π

* The theorem still holds when it is assumed only that functions X and Y are continuous.

and not zero. This result becomes obvious if we notice that arc tan $y/x = \varphi$, where φ is the polar angle of the point (x, y); on each complete circuit round the origin (the pole) this angle increases by 2π.

Example 1. There is no function of which $-y\,dx + x\,dy$ is the total differential, since

$$\frac{\partial X}{\partial y} = \frac{\partial}{\partial y}(-y) = -1, \quad \text{whilst} \quad \frac{\partial Y}{\partial x} = \frac{\partial}{\partial x}(x) = 1.$$

This fully agrees with the fact that $\int -y\,dx + x\,dy$ over any closed curve measures the absolute value of twice the area enclosed by the curve, so that it cannot be zero:

Example 2. Similarly, no primitive exists of the expression $(3xy - 2y^2 + 4y)\,dx + (2x^2 - 3xy + 4x)\,dy$, since $\partial X/\partial y = 3x - 4y + 4$, whilst $\partial Y/\partial x = 4x - 3y + 4$.

Example 3. The expression $(4x + 2y)\,dx + (2x - 6y)\,dy$ is the total differential of some function, since $\partial X/\partial y = 2$ and $\partial Y/\partial x = 2$.

REMARK. If the differential form $X\,dx + Y\,dy$ is not a total differential, the quantity

$$u = \int_L X\,dx + Y\,dy$$

is a function not only of point $P(x, y)$—the final point of the path of integration, but also of the entire path L joining the fixed point P_0 and point P. The expression $X\,dx + Y\,dy$ is, as before, the *differential* of the quantity u:

$$du = X\,dx + Y\,dy,$$

but only in the sense that, as P varies along a *given* curve L, it is the principal part of the increment Δu, linear in dx and dy. In this case dx and dy, instead of being arbitrary infinitesimal increments of variables x and y, are those obtained by the point $P(x, y)$ on curve L; the function u is in fact a function of a single independent variable.

Similar propositions hold in the spatial case.

THEOREM. **A necessary and sufficient condition for the differential form**

$$\boldsymbol{X(x, y, z)\,dx + Y(x, y, z)\,dy + Z(x, y, z)\,dz},$$

in which functions X, Y and Z have continuous partial derivatives in domain Ω, to be the total differential of a function $I(x, y, z)$ in Ω is that

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x}, \quad \frac{\partial X}{\partial z} = \frac{\partial Z}{\partial x}, \quad \frac{\partial Y}{\partial z} = \frac{\partial Z}{\partial y} \tag{A_1}$$

at every point of the domain.

The proof is omitted since it is similar to the above.

THEOREM. **Let functions $X(x, y, z)$, $Y(x, y, z)$ and $Z(x, y, z)$ be continuous along with their partial derivatives in a singly-connected domain Ω; then the necessary and sufficient condition for independence of the line integral**

$$\int_{P_0}^{P} X(x, y, z)\, dx + Y(x, y, z)\, dy + Z(x, y, z)\, dz$$

of the contour of integration in domain Ω (or what amounts to the same thing, its vanishing over any closed path) is that the integrand element $X\,dx + Y\,dy + Z\,dz$ be the total differential in domain Ω of some function $I(x, y, z)$:

$$X\,dx + Y\,dy + Z\,dz = dI.$$

187. Determination of the Primitive.

I. PLANE CASE. Let $X(x, y)\,dx + Y(x, y)\,dy$ be a total differential, i.e. the condition is satisfied:

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x}. \tag{A}$$

Since the integral of the expression:

$$I(x, y) = \int_{(x_0, y_0)}^{(x, y)} X\,dx + Y\,dy \tag{*}$$

leads to a function of two independent variables—the primitive—equation (A) is called *the equation of integrability*. The expression $X\,dx + Y\,dy$ has an infinite set of primitives, which differ only by a constant. In fact, if

$$du_1(x, y) = X\,dx + Y\,dy \quad \text{and} \quad du_2(x, y) = X\,dx + Y\,dy,$$

then

$$d[u_1(x, y) - u_2(x, y)] = 0,$$

whence it follows that $u_1(x, y) - u_2(x, y) = \text{const}$ (see Sec. 145). Let $u(x, y)$ denote a primitive of $X\,dx + Y\,dy$ in the singly-connected domain D. Since $I(x, y)$ is also a primitive by what has been proved, we have $I(x, y) = u(x, y) + C$; the fact that $I(x_0, y_0) = 0$ gives us $C = -u(x_0, y_0)$, so that

Thus
$$I(x, y) = u(x, y) - u(x_0, y_0).$$

$$\int_{(x_0, y_0)}^{(x, y)} X\,dx + Y\,dy = u(x, y) - u(x_0, y_0),$$

or
$$\int_{(x_0, y_0)}^{(x, y)} du(x, y) = u(x, y) - u(x_0, y_0);$$

the integrals are taken over any path joining the points $P_0(x_0, y_0)$ and $P(x, y)$.

We have obtained the *Newton–Leibniz formula for line integrals.*

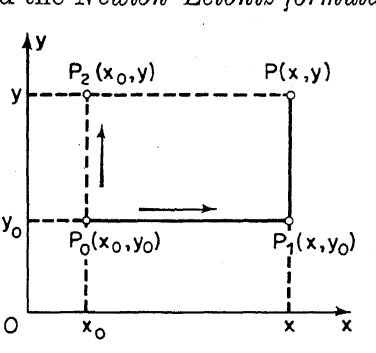

FIG. 63

Let us express (*) in terms of ordinary integrals. This is done most simply by taking as the contour of integration a step-line with steps parallel to the axes, provided it belongs to domain D.

Suppose the contour of integration is the step-line P_0P_1P (Fig. 63). We now have $y = y_0$ and $dy = 0$ along P_0P_1, and $dx = 0$ along P_1P. We get:

$$\int_{(x_0, y_0)}^{(x, y)} X(x, y)\,dx + Y(x, y)\,dy = \int_{x_0}^{x} X(x, y_0)\,dx + \int_{y_0}^{y} Y(x, y)\,dy.$$

Similarly, integration along P_0P_2P gives

$$\int_{(x_0, y_0)}^{(x, y)} X(x, y)\,dx + Y(x, y)\,dy = \int_{x_0}^{x} X(x, y)\,dx + \int_{y_0}^{y} Y(x_0, y)\,dy.$$

A function is in fact generally found from its total differential by means of these formulae.

The general expression for the primitive has the form

$$\int_{(x_0, y_0)}^{(x, y)} X\,dx + Y\,dy + C = \begin{cases} \int_{x_0}^{x} X(x, y_0)\,dx + \int_{y_0}^{y} Y(x, y)\,dy + C, \\ \int_{x_0}^{x} X(x, y)\,dx + \int_{y_0}^{y} Y(x_0, y)\,dy + C, \end{cases}$$

where $P_0(x_0, y_0)$ is any given point and C is an arbitrary constant.

This is similar to the formula for the ordinary indefinite integral (Sec. 94).

Example. Let us find the primitive of

$$(4x + 2y)\,dx + (2x - 6y)\,dy$$

(Sec. 186, Example 3), vanishing at the point (0, 0). The above formulae yield:

$$\int_{(0,0)}^{(x, y)} (4x + 2y)\,dx + (2x - 6y)\,dy = \int_0^x 4x\,dx + \int_0^y (2x - 6y)\,dy = 2x^2 + 2xy - 3y^2$$

or

$$\int_{(0,0)}^{(x, y)} (4x + 2y)\,dx + (2x - 6y)\,dy = \int_0^x (4x + 2y)\,dx + \int_0^y -6y\,dy = 2x^2 + 2xy - 3y^2.$$

II. Spatial case. The equations

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x}, \quad \frac{\partial X}{\partial z} = \frac{\partial Z}{\partial x}, \quad \frac{\partial Y}{\partial z} = \frac{\partial Z}{\partial y} \qquad (A_1)$$

are called the *integrability conditions*. Suppose that they hold in a singly-connected domain of space Ω; then $X(x, y, z)\,dx + Y(x, y, z)\,dy + Z(x, y, z)\,dz$ is the total differential of a function which is equal in domain Ω (apart from an arbitrary constant) to the integral of the total differential over any contour joining the fixed point $P_0(x_0, y_0, z_0)$ to the variable point $P(x, y, z)$:

$$\int_{(x_0, y_0, z_0)}^{(x, y, z)} X\,dx + Y\,dy + Z\,dz = \int_{(x_0, y_0, z_0)}^{(x, y, z)} du(x, y, z) = u(x, y, z) - u(x_0, y_0, z_0)$$

(Newton–Leibniz formula).

The function is found in practice from its total differential as in the plane case. The most convenient formulae are obtained by integration along a step-line joining the initial point $P_0(x_0, y_0, z_0)$ to the final point $P(x, y, z)$, the steps being parallel to the axes and the line being contained in domain Ω.

Let us take, for instance, step-line $P_0P_1P_2P$ (Fig. 64) as contour of integration. We have along P_0P_1: $y = y_0$, $z = z_0$ and

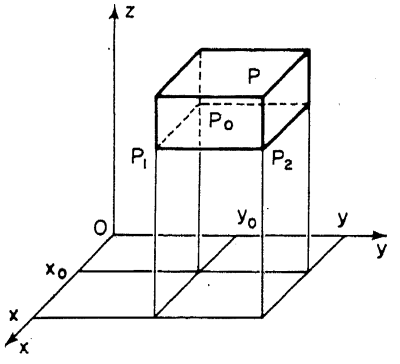

FIG. 64

$dy = 0$, $dz = 0$; along P_1P_2 we have $z = z_0$ and $dx = 0$, $dz = 0$; finally, along P_2P we have $dx = 0$, $dy = 0$. Hence

$$u(x,y,z) - u(x_0, y_0, z_0) = \int_{(x_0,y_0,z_0)}^{(x,y,z)} X\,dx + Y\,dy + Z\,dz$$

$$= \int_{x_0}^{x} X\,(x, y_0, z_0)\,dx + \int_{y_0}^{y} Y\,(x, y, z_0)\,dy + \int_{z_0}^{z} Z(x, y, z)\,dz.$$

We arrive at a similar formula if we integrate along another rib of the parallelepiped P_0P, leading from point P_0 to point P.

The general expression for the primitive can be written as say

$$\int_{(x_0,y_0,z_0)}^{(x,y,z)} X\,dx + Y\,dy + Z\,dz + C$$

$$= \int_{x_0}^{x} X(x, y_0, z_0)\,dx + \int_{y_0}^{y} Y(y, y, z_0)\,dy + \int_{z_0}^{z} Z(x, y, z)\,dz + C,$$

where $P_0(x_0, y_0, z_0)$ is any given point of the domain and C is an arbitrary constant.

188. General Approach to the Solution of Problems. Problems of Hydrodynamics and Thermodynamics. Suppose that a magnitude u

can be regarded as a function of the arc s of a given plane curve L, $u = u(s)$, possessing the property of additiveness, i.e.

$$u(s) = u(s_1) + u(s_2) + \cdots + u(s_n),$$

if $s = s_1 + s_2 + \cdots + s_n$. We take a piece of curve L, as small as desired, at an arbitrary point $P(x, y)$, and characterized by the increments dx and dy, and we find the element (differential) du of u corresponding to this piece.

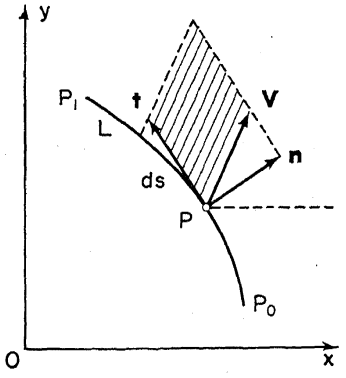

FIG. 65

It is given in the cases more commonly encountered as

$$du = X\,dx + Y\,dy, \qquad (*)$$

where X and Y are known functions of the point P. Summation over all the elements along curve L now gives u:

$$u = \int_L X\,dx + Y\,dy.$$

If expression (*) for du is a total differential, the integral can be taken along any curve and u becomes simply a *function of a point* (the final point of the path of integration). Otherwise, u depends essentially on the path of integration.

We shall take as examples some important problems of hydrodynamics and thermodynamics.

I. A PROBLEM OF HYDRODYNAMICS. A fluid flow is said to be stationary if the velocity of its particles at a given point is independent of time. To say that the fluid is *incompressible* implies that its density is constant; this means mathematically that the mass of fluid is measured by the volume that it occupies. A vector $\mathbf{V}(P)$, the particle velocity, is associated with each point $P(x, y, z)$ of the domain in which the stationary fluid flow occurs; we write $X = X(P)$, $Y = Y(P)$, $Z = Z(P)$ for the projections of the velocity vector on Ox, Oy, Oz; these functions along with their partial derivatives will be assumed continuous. However, we shall only consider plane flow. The flow is plane when all the fluid particles move parallel to a plane, the velocities of particles on a straight line perpendicular to this plane being the same. If we take this plane to be Oxy, we only need to consider the flow diagram in Oxy since it is the same in all parallel planes.

We find the quantity of fluid K_1 flowing past a given curve L in unit time (Fig. 65). We take an infinitesimal arc ds of L at an arbitrary point P and let ΔK_1 be the quantity of fluid traversing ds to some definite side of the curve per unit time (the side towards which the normal $\mathbf{n}$ is directed in Fig. 65).

The principal part of ΔK_1, i.e. the element dK_1 of the quantity of fluid, is found on the assumption that ds is directed along the tangential vector

$\mathbf{t}$ and that the velocity vector $\mathbf{V}$ does not change as we move along ds. Vector $\mathbf{V}$ indicates the path traversed by a particle in unit time, and the quantity of fluid dK_1 is measured with the present conditions by the area of the parallelogram constructed on vectors $\mathbf{V}$ and $\mathbf{t}$. But this area is equal to the product of ds and the height of the parallelogram, i.e. the projection of $\mathbf{V}$ on the normal $\mathbf{n}$:

$$dK_1 = \mathrm{proj}_n \mathbf{V}\, ds = |\mathbf{V}| \cos(\mathbf{V}, \mathbf{n})\, ds,$$

where $(\mathbf{V}, \mathbf{n})$ is the angle between $\mathbf{V}$ and $\mathbf{n}$.

We know that the cosine of the angle between two directions is given by

$$\cos(\mathbf{V}, \mathbf{n}) = \cos(\mathbf{V}, x) \cos(\mathbf{n}, x) + \cos(\mathbf{V}, y) \cos(\mathbf{n}, y),$$

and since

$$\cos(\mathbf{n}, x) = \cos(\mathbf{t}, y) \quad \text{and} \quad \cos(\mathbf{n}, y) = -\cos(\mathbf{t}, x),$$

we have

$$\cos(\mathbf{V}, \mathbf{n}) = \cos(\mathbf{V}, x) \cos(\mathbf{t}, y) - \cos(\mathbf{V}, y) \cos(\mathbf{t}, x).$$

Therefore

$$dK_1 = |\mathbf{V}| \cos(\mathbf{V}, x) \cos(\mathbf{t}, y)\, ds - |\mathbf{V}| \cos(\mathbf{V}, y) \cos(\mathbf{t}, x)\, ds,$$

i.e.*

$$dK_1 = X(x, y)\, dy - Y(x, y)\, dx.$$

Integration over the whole of L gives

$$K_1 = \int_L X\, dy - Y\, dx.$$

We call K_1 the flux across curve L; it is equal to the quantity of fluid traversing L in unit time.

Let L be a closed curve. Integral (*) now gives the total quantity of fluid traversing L in unit time in the direction of the *outward normal* (if the integration is in a positive sense, i.e. anticlockwise).

Let there be neither positive nor negative sources in the domain D bounded by contour L (i.e. there are no points at which fluid either enters or is absorbed). It is clear from the physical point of view that in this case, however much fluid enters domain D, the same amount must flow out; hence integral (*), expressing the gain or loss of fluid, must be equal to zero.

The plane stationary flow of an incompressible fluid without sources is thus characterized by the fact that

$$\int_L X\, dy - Y\, dx = 0$$

for every closed curve L; and as we know, this condition is equivalent to $\partial X/\partial x = -\partial Y/\partial y$, i.e. $\partial X/\partial x + \partial Y/\partial y = 0$.

Integral (*) does not depend on the contour of integration; the primitive

$$v(P) = v(x, y) = \int_{(P_0)}^{(P)} X\, dy - Y\, dx$$

* This formula can easily be deduced vectorially, from the fact that the area of the parallelogram constructed on vectors $\mathbf{V}$ and $\mathbf{t}$ is equal to the absolute value of the vector product $\mathbf{V} \times \mathbf{t}$.

is called the *stream function*. The stream function has a simple physical significance: the difference $v(P_1) - v(P_0)$ between its values at two points P_1 and P_0 measures the quantity of fluid traversing an arbitrary curve between these points.

In addition to integral (*) a further important characteristic may be introduced for the plane stationary flow of an incompressible fluid. This is the quantity K_2, where the element dK_2 along curve L is the product of ds and the projection of vector $\mathbf{V}$ on tangential vector $\mathbf{t}$ (see Fig. 65):

$$dK_2 = |\mathbf{V}| \cos(\mathbf{V}, \mathbf{t})\, ds = X\, dx + Y\, dy.$$

We have

$$K_2 = \int_L X\, dx + Y\, dy. \qquad (**)$$

K_2 is the quantity of fluid travelling *along* curve L per unit time; it is called the *circulation* along L.

Suppose that the circulation along any closed contour vanishes, i.e. that integral (**) is independent of the path of integration. The flow in this case is said to be *irrotational* (without vortices).

Thus *the plane stationary irrotational flow of an incompressible fluid is characterized by the fact that*

$$\int_L X\, dx + Y\, dy = 0$$

for any closed contour L; as we know, this condition is equivalent to $\partial X/\partial y = \partial Y/\partial x$, i.e. $\partial Y/\partial x - \partial X/\partial y = 0$.

The primitive

$$u(P) = u(x, y) = \int_{(P_0)}^{(P)} X\, dx + Y\, dy$$

is called the *velocity potential*.

Finally, *let both sources and vortices be absent* in the plane stationary flow of an incompressible fluid. The stream function $v(x, y)$ and velocity potential $u(x, y)$ are now closely related. A comparison of the two cases yields the so-called *Cauchy–Riemann* conditions:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} (= X), \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} (= Y).$$

A single complex function, $u + iv$, called the *characteristic function*, is formed from functions u and v. A number of problems of great importance in hydrodynamics can be solved with its aid.

II. Problems of thermodynamics. When we speak of the state of a body we imply a system of magnitudes indicating its physical condition. In thermodynamics these magnitudes are generally: the *pressure* p, *volume* v and *absolute temperature* T. In other words, the state of a body is specified by the three quantities p, v, T. But these are connected by an equation, called the equation of state. Hence the state of a body is in fact defined by two quantities, say p and v (the third, T, is a function of p and v).

Geometrically, for each *state* there is a corresponding *point* $P(p, v)$ of the Opv plane, i.e. for each *process* (consisting of a sequence of changes of

state) a *curve*. It is called the *diagram of the process*. When the body returns to its initial state the process is said to be *cyclical*, and its diagram is a *closed* curve.

Let us take the elementary case when the body is an ideal gas, i.e. a gas subject to the Mendeleev–Clapeyron equation of state:

$$pv = RT \quad (R = \text{const}).$$

We pose the problem of finding the quantity of heat Q absorbed by the gas during the process described by a curve L. We take an arbitrary piece of diagram L from a point $P(p, v)$ to $P(p + dp, v + dv)$, describing an "infinitesimal process"; let ΔQ denote the corresponding quantity of heat. This heat implies an increment in the mechanical energy of the gas particles, i.e. in the last analysis an increment in the temperature dT and a certain amount of work in changing the volume by dv. We find the element dQ (the principal part of ΔQ, linear in dT and dv) by supposing that, on the basis of the principle of superposition of small operations, the heat absorbed is the sum of two terms: (1) the heat expended in raising the temperature by dT at constant volume v; (2) the heat expended in the work of expansion dv at constant temperature T. The first is equal to $c_v dT$, where c_v is the heat capacity of the gas at constant volume. The second is given by the product of the pressure p and the volume dv.

Thus

$$dQ = c_v dT + ap\, dv,$$

where $a = 1/427$ cal/kg is the thermal equivalent of work. (Multiplication by a reduces the mechanical work pdv to thermal units of work.)

In accordance with the Mendeleev-Clapeyron equation:

$$dT = \frac{v}{R} dp + \frac{p}{R} dv;$$

we obtain on introducing this into the expression for dQ:

$$dQ = \frac{c_v}{R} v\, dp + \frac{c_v + aR}{R} p\, dv.$$

The significance of $c_v + aR$ is as follows. We find from the last two equations with constant pressure p, $dp = 0$:

$$dT = \frac{p}{R} dv, \quad dQ = (c_v + aR) \frac{p}{R} dv,$$

i.e.

$$dQ = (c_v + aR)\, dT,$$

whence it is clear that the coefficient of dT is simply the heat capacity c_p at constant pressure:

$$c_v + aR = c_p \qquad (***)$$

(c_p and c_v are assumed *constant*).

Hence, finally:

$$dQ = \frac{c_v}{R} v\, dp + \frac{c_p}{R} p\, dv.$$

We find the required quantity of heat Q simply by integrating the expression for dQ over the process diagram L:

$$Q = \int_L \frac{c_v}{R} v\,dp + \frac{c_p}{R} p\,dv.$$

The condition for independence of the integral on the path of integration is not fulfilled here. For:

$$\frac{\partial}{\partial v}\left(\frac{c_v}{R} v\right) = \frac{c_v}{R}, \quad \frac{\partial}{\partial p}\left(\frac{c_v}{R} p\right) = \frac{c_p}{R},$$

whilst $c_v \neq c_p$ in view of relationship (***). Hence Q is essentially dependent on the contour L, i.e. Q is not a function of the point $P(p, v)$. Our discussion shows that *the amount of heat absorbed is not a function of the gas state;* it depends not only on the final state, but also on *how* the gas has arrived at the final state, in other words, on the system of all intermediate states. In particular. a cyclical process (cycle) in general implies the absorption or subtraction of heat, which is used in the production of external work.

A quantity which characterizes a process and is of great importance in thermodynamics is the entropy S. The entropy S is defined as the additive function such that the element dS corresponding to the piece of the process diagram from point $P(p, v)$ to $P(p + dp, v + dv)$ is equal to the element of heat dQ divided by the temperature T:

$$dS = \frac{dQ}{T},$$

whence

$$S = \int_L \frac{dQ}{T},$$

where L is the diagram.

We have for an ideal gas:

$$S = \int_L \frac{dQ}{T} = \int_L \frac{c_v}{RT} v\,dp + \frac{c_p}{RT} p\,dv,$$

and since $RT = pv$,

$$S = \int_L \frac{c_v}{p}\,dp + \frac{c_p}{v}\,dv.$$

The integrand element is now a total differential:

$$\frac{\partial}{\partial v}\left(\frac{c_v}{p}\right) = \frac{\partial}{\partial p}\left(\frac{c_v}{v}\right) = 0$$

i.e. *the entropy is a function of the gas state; its magnitude is not dependent on how the gas passes from its initial to its final state, and in particular, it vanishes for any cycle.*

Integration gives us:

$$S = \ln C p^{c_v} v^{c_p}.$$

We take two particular cases:

(1) For an *isothermal* process ($T = \text{const.}$, $dT = 0$), $dQ = ap\,dv$ and

$$Q = a\int_L p\,dv.$$

It follows that, when L is a closed curve, the amount of absorbed heat Q is proportional to the area bounded by the diagram. (The diagram of an isothermal process is called an isotherm.)

(2) If the process is adiabatic ($Q = \text{const}$, $dQ = 0$, i.e. also $dS = 0$, i.e. $S = \text{const.}$), we have from the expression for S:

$$p^{c_v}v^{c_p} = \text{const.},$$

whence

$$pv^k = \text{const.},$$

where $k = c_p/c_v > 1$. This is the equation of the diagram of an adiabatic process (called an adiabat) in an ideal gas. Adiabats are *polytropic curves.*

The following point is worth mentioning: although the expression $dQ = c_v v\,dp/R + c_p p\,dv/R$ is not a total differential, it becomes one after multiplication by a certain function ($1/T = R/pv$). This situation is discussed from the general point of view in Sec. 195.

3. Surface Integrals

189. Integrals over a Surface Area and Co-ordinate Surface Integrals.

I. INTEGRALS OVER A SURFACE AREA. The integral over a surface area is a generalization of the double integral, just as the line integral over an arc is a generalization of the ordinary integral.

Suppose we are given (1) a continuous surface S satisfying the conditions for measurability (Sec. 177); (2) a continuous function $f(P)$ of a point P on surface S.

We subdivide S into n pieces $q_1, q_2, \ldots, q_n$ with areas $\Delta q_1, \Delta q_2, \ldots, \Delta q_n$ respectively; we take arbitrary points $P_1, P_2, \ldots, P_n$ on the surfaces and let $f(P_1), f(P_2), \ldots, f(P_n)$ be the values of $f(P)$ at these points.

We form the sum

$$I_n = \sum_{i=1}^{n} f(P_i)\,\Delta q_i,$$

called the n-th *integral sum (with respect to area) for function* $f(P)$ *and surface* S.

Definition. **The limit I of integral sum I_n as the greatest diameter of a piece of surface tends to zero is called the *integral over the surface* S of the function $f(P)$ and is written as**

$$I = \iint_S f(P)\,dq,$$

where dq is an elementary area of the surface (or simply, an element of) S.

There is no need to dwell on the notation and terminology, since this is precisely the same as in previous cases.

The general properties of multiple integrals can of course be applied directly to surface integrals. In particular, if surface S is divided into a number of parts $q_1, q_2, \ldots, q_n$, then

$$\iint\limits_S f(P)\,dq = \iint\limits_{q_1} f(P)\,dq + \iint\limits_{q_2} f(P)\,dq + \cdots + \iint\limits_{q_n} f(P)\,dq.$$

We turn to the evaluation of surface integrals.

Let surface S be cut by any straight line parallel to Oz in not more than one point. The equation of the surface can now be written as $z = z(x, y)$, where $z(x, y)$ is a *single-valued* function of the orthogonal projection $P_1(x, y)$ of point $P(x, y, z)$ of surface S on Oxy.

If $f(P)$ is given in terms of the co-ordinates of point P, i.e. in the form $f(x, y, z)$, substitution of $z(x, y)$ for z in this expression gives us a function of two variables $f[x, y, z(x, y)]$. Moreover (see Sec. 177)

$$dq = \sqrt{1 + z'^2_x(x, y) + z'^2_y(x, y)}\,dx\,dy.$$

Therefore

$$\iint\limits_S f(x, y, z)\,dq = \iint\limits_D f[x, y, z(x, y)]\sqrt{1 + z'^2_x + z'^2_y}\,dx\,dy,$$

where D is the projection of surface S on Oxy. We have obtained a double integral. It is not necessary to give a special statement of the rule for evaluation of a surface integral.

When the domain S of integration does not satisfy the condition mentioned, it has to be subdivided so that each part separately satisfies the condition. The surface integral can then be written as the sum of the double integrals over domains lying in the co-ordinate planes.

Thus *evaluation of an integral over any surface reduces to evaluation of double integrals.*

II. Co-ordinate surface integral. This type of integral is more useful than the integral over a surface.

We must dwell on the question of the *orientation of a surface* before defining the integral. Whereas the sense of a smooth curve is decided with the aid of the tangent*, *the sense of a surface is determined with the aid of the normal.*

* A curve has an orientation (sense) if we indicate one of the two possible directions of the tangent, assuming that this direction changes continuously with the point of contact.

Definition. **A surface is said to be *orientated* if we indicate one of the two possible directions of the normal to it, on the assumption that this direction varies continuously with the point of the surface at which the normal is drawn.**

Suppose we have a directed normal at any given point P of a surface S. We displace the point along with its normal from a point P_0 along an arbitrary path on the surface so that the normal direction changes *continuously*, i.e. the directions at infinitesimally neighbouring points are infinitesimally close to each other.

If, when point P returns to its initial position P_0, the direction of the normal is the same as the original direction, surface S is said to be *two-sided*. Otherwise, it is said to be *one-sided*.

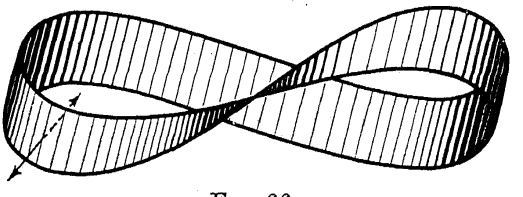

FIG. 66

The *Möbius band* is an example of a one-sided surface. A model of it is easily made from a straight strip of paper. The strip is bent round as when forming the model of a cylinder, except that we first give the paper a twist (Fig. 66). If we start from any given point of the Möbius band with a definite direction of the normal, we can return to the same point with the opposite normal direction whilst preserving continuity and never crossing the edge of the band.

We shall be concerned exclusively with *two-sided surfaces*.

The choice of normal direction at any point of a two-sided surface gives it an orientation by distinguishing one side from the other. An orientated two-sided surface is *one on which a definite side has been chosen*.

In the case of a closed surface its outer side, i.e. the side towards external space (with respect to the domain bounded by the surface) is usually called the *positive side*.

On projecting a surface element dq on to the co-ordinate plane we obtain a corresponding element $d\sigma$ of the plane. We supplement what has been said about orientation by making the following agreement:

The projection $d\sigma$ (say on Oxy) **of element dq of the surface will be taken with the + or — sign depending on whether the normal di-**

rected towards the chosen side of the surface forms an acute or obtuse angle with the perpendicular to the plane of projection (say with Oz).

The reader will soon see the point of our agreement. It will be clear that, by orientating elements of the co-ordinate plane itself (Oxy), the agreement leads to a similar state of affairs in the two-dimensional case as holds in one dimension*.

We take the above stipulations into account when defining the co-ordinate surface integral, this latter being similar to the line integral.

The process of forming a co-ordinate surface integral differs from the formation of the integral over a surface area only in multiplying the value of function $f(P)$ (see I) at point P of surface S by the orientated element $d\sigma$ of the co-ordinate plane Oxy (or Oxz or Oyz) corresponding to the surface element dq instead of by dq itself.

Thus the integral sum has the form

$$I_n = \sum_{i=1}^{n} f(P_i)\, \Delta\sigma_i, \qquad (*)$$

whilst its limit is written as

$$I = \iint_S f(P)\, d\sigma = \iint_S f(P)\, dx\, dy, \qquad (**)$$

the sign of element $d\sigma = dx\, dy$ being in accordance with our agreement.

The sum (*) is called the n-th *integral sum with respect to the xy co-ordinate plane for function $f(P)$ and surface S*.

Definition. **The limit I (**) of integral sum I_n (*) when the greatest diameter of a sub-domain tends to zero is called the xy co-ordinate surface integral of function $f(P)$ over surface S.**

We can similarly define

$$I = \iint_S f(P)\, dx\, dz, \quad I = \iint_S f(P)\, dy\, dz.$$

The notation for the surface integral with respect to a co-ordinate plane must be accompanied by an indication of the direction of integration, i.e. by a choice of side of surface S.

Co-ordinate surface integrals have the same basic properties as ordinary surface integrals and multiple integrals in general. In

* The projection (dx or dy) of an elementary arc ds is automatically orientated: it is positive if the tangent (in the plane of the curve!) forms an acute angle with the axis of projection, and negative otherwise.

particular, if surface S is divided into parts $q_1, q_2, \ldots, q_n$ (with the same orientation), then

$$\iint_S f(P)\,d\sigma = \iint_{q_1} f(P)\,d\sigma + \iint_{q_2} f(P)\,d\sigma + \cdots + \iint_{q_n} f(P)\,d\sigma.$$

Moreover, we have the following property in connection with an orientated surface:

If the surface orientation (side) is changed, the integral takes the opposite sign:

$$\iint_{+S} f(P)\,d\sigma = -\iint_{-S} f(P)\,d\sigma,$$

where $+S$ *and* $-S$ *denote the two sides of surface* S.

In fact, changing the side of surface S implies reversing the normal direction, i.e. changing the signs of all the $\Delta\sigma$ in the integral sums.

This property is similar to the corresponding one for line integrals and leads to an important property of surface integrals.

THEOREM. **If a domain Ω of space bounded by a closed surface S is sub-divided with the aid of closed surfaces $S_1, S_2, \ldots, S_n$, the surface integral over the whole of S** (say over the positive side) **is equal to the sum of the integrals over the corresponding (positive) sides of the orientated surfaces $S_1, S_2, \ldots, S_n$:**

$$\iint_S f(P)\,d\sigma = \iint_{S_1} f(P)\,d\sigma + \iint_{S_2} f(P)\,d\sigma + \cdots + \iint_{S_n} f(P)\,d\sigma.$$

(We naturally assume that the integrand is defined and continuous throughout Ω.)

Proof. On the right-hand side each surface lying inside domain Ω is altogether covered twice during integration, over both sides; the sum of the corresponding integrals vanishes, and as a result there remain in the sum giving the integral over the whole of S only the integrals over the parts of the boundary of Ω.

This is what we needed to prove.

Surface integrals with respect to co-ordinate planes are evaluated by transforming them to double integrals. Given

$$\iint_S f(x, y, z)\,dx\,dy,$$

let the surface S be cut by a straight line parallel to Oz in not more than one point. The element $d\sigma = dx\,dy$ now has a constant

sign (*plus* if the integration is over the *outside* of the surface, i.e. in the direction of increasing z, and *minus* if over the *inside*). Let the equation of the surface be $z = z(x, y)$, where $z(x, y)$ is a single-valued function, and let the integration be over the outside of S. Then

$$\iint_S f(x, y, z)\, dx\, dy = \iint_D f[x, y, z(x, y)]\, dx\, dy,$$

where D is the projection of surface S on Oxy. In the general case the surface has to be subdivided so that the integral over each piece can be evaluated by the method described.

Surface integrals with respect to the other co-ordinate planes are evaluated in the same way.

The rule for evaluating a surface integral will now be obvious and may be omitted.

Example. Find

$$I = \iint_S xyz\, dx\, dy,$$

taken over the outside of the sphere $x^2 + y^2 + z^2 = 1$ contained in the first and eighth octants: $x \geqslant 0$, $y \geqslant 0$.

The z co-ordinate cannot be expressed as a single-valued function of x and y for the whole of the surface of integration S. We divide S into two parts: S_2, lying above Oxy, and S_1, lying below. Their equations are respectively: $z_2 = \sqrt{1 - x^2 - y^2}$ and $z_1 = -\sqrt{1 - x^2 - y^2}$. We have:

$$I = \iint_S xyz\, dx\, dy = \iint_{S_2} xyz\, dx\, dy + \iint_{S_1} xyz\, dx\, dy,$$

and since the second integral on the right-hand side is taken over the lower side S_1 of the part of the sphere contained in the eighth octant, we have

$$I = \iint_{S_2} xyz\, dx\, dy - \iint_{-S_1} xyz\, dx\, dy,$$

where $-S_1$ is the upper side of this part.

Transformation from surface to double integrals gives

$$I = \iint_D xy\sqrt{1 - x^2 - y^2}\, dx\, dy - \iint_D -xy\sqrt{1 - x^2 - y^2}\, dx\, dy$$

$$= 2\iint_D xy\sqrt{1 - x^2 - y^2}\, dx\, dy,$$

where D is the quadrant of the circle $x^2 + y^2 \leqslant 1$ lying in the first co-ordinate quadrant of Oxy, D being the projection of both S_1 and S_2.

The double integral is evaluated by using polar instead of Cartesian co-ordinates:

$$I = 2 \iint_D \varrho^2 \sin\varphi \cos\varphi \sqrt{1 - \varrho^2}\ \varrho\, d\varrho\, d\varphi$$

$$= \int_0^{\frac{\pi}{2}} \sin 2\varphi\, d\varphi \int_0^1 \varrho^3 \sqrt{1 - \varrho^2}\, d\varrho = 1 \cdot \frac{2}{15} = \frac{2}{15}.$$

If the integrand is defined and continuous in a domain and not merely on a single surface, the surface integral (or co-ordinate surface integral) over an arbitrary surface in the domain is a "functional", which depends on the surface.

190. Component Surface Integrals. Stokes' Formula.

I. The combination (sum) of surface integrals:

$$\iint_S X\, dy\, dz + \iint_S Y\, dx\, dz + \iint_S Z\, dx\, dy,$$

where $X = X(x, y, z)$, $Y = Y(x, y, z)$, $Z = Z(x, y, z)$ are functions of variables x, y, z, is to be encountered in analysis and its applications: it is written with a single surface integral symbol:

$$\iint_S X\, dy\, dz + Y\, dx\, dz + Z\, dx\, dy.$$

An important problem can be posed for this "composite" surface integral, similar to that already discussed for composite line integrals: *what conditions must functions X, Y, Z satisfy for the surface integral over any closed surface to vanish, or what amounts to the same thing, for the surface integral to depend only on the boundary of the surface.*

We shall leave the solution of this problem until Ostrogradskii's formula has been obtained (Sec. 191), from which the solution follows at once. We turn now to Stokes' formula for composite surface integrals.

II. Stokes' formula* is a direct generalization of Green's formula (Sec. 184).

* G. Stokes (1819–1903).

STOKES' THEOREM. **If functions $X(x, y, z)$, $Y(x, y, z)$ $Z(x, y, z)$ are continuous together with their first order partial derivatives in a domain Ω containing a continuous and measurable surface S, we have**

$$\iint_S \left(\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z}\right) dy\,dz + \left(\frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x}\right) dx\,dz + \\ + \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right) dx\,dy = \int_L X dx + Y dy + Z dz, \qquad (*)$$

where L is the boundary of surface S.

The directions of integration over curve L and surface S are made to agree with the aid of the following visual rule: if a man moves on the side of surface S over which the integration is carried out along the boundary L in the direction of the line integration, surface S must remain on his left.

Expression (*) *is known as* Stokes' formula.

Proof. We suppose first that surface S is cut by any straight line parallel to Oz in not more than one point. Let $z = z(x, y)$ be the equation of S.

We transform the composite surface integral

$$I = \iint_S \frac{\partial X}{\partial y}\, dx\, dy - \frac{\partial X}{\partial z}\, dx\, dz.$$

Let the integral be taken over the upper side of the surface.

On writing $d\sigma_{xy}$ and $d\sigma_{xz}$ for the projections of surface element dq on Oxy and Oxz respectively, we have:

$$d\sigma_{xy} = \cos\gamma\, dq, \quad d\sigma_{xz} = \cos\beta\, dq,$$

where γ and β are the angles of orientation of the normal to S to Oz and Oy respectively. Hence

$$d\sigma_{xz} = \frac{\cos\beta}{\cos\gamma}\, d\sigma_{xy},$$

and since (see Sec. 166) $\cos\beta/\cos\gamma = -z'_y$, we have

$$d\sigma_{xz} = -z'_y\, d\sigma_{xy}, \quad \text{i.e.} \quad dx\,dz = -z'_y\, dx\,dy,$$

so that

$$I = \iint_S \left(\frac{\partial X}{\partial y} + \frac{\partial X}{\partial z} z'_y\right) dx\,dy.$$

We reduce this surface integral to a double integral. To do this, we have to replace variable z in the integrand by its expression in terms of x and y from the equation of the surface $z = z(x, y)$. But the integrand now becomes equal to the partial derivative with respect to y of the function of a function got from $X(x, y, z)$ after substituting $z(x, y)$ for z. In fact,

$$\frac{\partial}{\partial y} X[x, y, z(x, y)] = \frac{\partial X}{\partial y} + \frac{\partial X}{\partial z} z'_y .$$

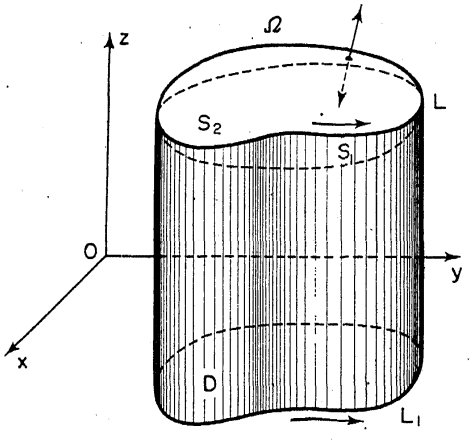

FIG. 67

Hence we have, on setting $X[x, y, z(x, y)] = X_1(x, y)$:

$$I = \iint_D \frac{\partial X_1}{\partial y} dx\, dy ,$$

where D is the projection of surface S on Oxy (Fig. 67). We obtain on applying Green's formula (Sec. 184):

$$I = \iint_D \frac{\partial X_1}{\partial y} dx\, dy = - \int_{L_1} X_1\, dx ,$$

where L_1 is the boundary of D, which is evidently the projection of the spatial curve L on Oxy, and integration is in the positive direction round L_1.

The line integral on the right-hand side of the last equation is easily seen to be equal to the integral of $X(x, y, z)$ over the spatial curve L in the direction corresponding to the positive direction round its projection L_1 (both directions are shown by arrows in Fig. 67).

Thus

$$I = \iint_S \frac{\partial X}{\partial y} dx\, dy - \frac{\partial X}{\partial z} dx\, dz = -\int_L X\, dx. \qquad (A)$$

To preserve this formula, a change of the side of S over which the integration is carried out (upper to lower) involves changing the direction of integration round curve L.

This shows that the directions of integration are in agreement with the rule given above.

We note further that (A) holds for *any* continuous and measurable surface, and not merely for a surface which is cut by a straight line parallel to Oz in not more than one point. This can be proved in precisely the same way as in all the similar cases considered earlier.

A similar proof can also be given of the two relationships:

$$\iint_S \frac{\partial Y}{\partial z} dy\, dz - \frac{\partial Y}{\partial x} dy\, dx = -\int_L Y\, dy, \qquad (B)$$

$$\iint_S \frac{\partial Z}{\partial x} dz\, dx - \frac{\partial Z}{\partial y} dz\, dy = -\int_L Z\, dz. \qquad (C)$$

On adding formulae (A), (B), (C) term by term, we arrive at Stokes' formula (*).

The theorem is proved.

It may be noticed that the third term $(\partial Y/\partial x - \partial X/\partial y)\, dx\, dy$ after the symbol of the surface integral in formula (*) is the same as in Green's formula, whilst the other two terms are obtained from it by cyclical permutation of the letters x, y, z and X, Y, Z.

Moreover the same mnemonic can be used here: after the surface integral symbol the capitals (in the numerators) are arranged in the sequence $ZYXZYX$, and the small letters x, y, z (in the denominators) repeat the capitals in each bracket but in the reverse order.

Obviously, Stokes' formula reduces to Green's formula with $z = \text{const.}$ (i.e. when S is a plane).

Stokes' formula enables a surface integral, i.e. an integral over an orientated surface, to be replaced by a line integral over the surface boundary with corresponding orientation, and conversely, a line integral over a closed spatial curve can be replaced by an integral over some surface "stretched" over the contour of integration.

III. In Sec. 185 we proved the fundamental theorem on the independence of a line integral in a plane on the path of integration

with the aid of Green's formula; in the same way we can use Stokes' theorem to give a proof (omitted in Sec. 185) of the fundamental theorem for line integrals over a spatial curve.

For, let equation (A_1)

$$\frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x}, \quad \frac{\partial X}{\partial z} = \frac{\partial Z}{\partial x}, \quad \frac{\partial Y}{\partial z} = \frac{\partial Z}{\partial y} \tag{A_1}$$

be satisfied identically in a *singly-connected** domain Ω; on stretching a surface S in Ω over the closed curve L, we can conclude by Stokes' theorem that line integral (*) vanishes, no matter what the closed curve L.

Conversely, let line integral (*) vanish for any closed curve and let equations (A_1) be at the same time untrue at some point $P_0(x_0, y_0, z_0)$ of domain Ω, say

$$\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y} = \mu > 0 .$$

We draw the plane $z = z_0$ through P_0 parallel to Oxy. A δ-neighbourhood of P_0 can be taken in this plane such that $\partial Y/\partial x - \partial X/\partial y \geqslant \mu - \varepsilon > 0$ at every point of the neighbourhood. Bearing in mind that $dz = 0$ in domain δ, we conclude as in Sec. 185 that there exists a closed path in Ω over which integral (*) does not vanish, which contradicts our hypothesis. This proves the theorem and fills the the gap left over from Sec. 185.

191. Ostrogradskii's Formula.

I. Ostrogradskii's formula is so to speak an extension of Green's formula (Sec. 184) to the case of space; it connects a triple and surface integral.

OSTROGRADSKII'S THEOREM. **If functions $X(x, y, z)$, $Y(x, y, z)$, $Z(x, y, z)$ are continuous together with their first order partial derivatives in domain Ω, we have**

$$\iiint_{\Omega} \left(\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z}\right) dx\, dy\, dz = \iint_{S} X\, dy\, dz + Y\, dx\, dz + Z\, dx\, dy, \tag{*}$$

where S is the boundary of Ω and the integration is over the outside of S.

* With a domain of this type a surface lying wholly in the domain can be stretched over any closed curve.

Equation (*) is known as *Ostrogradskii's formula.*

Proof. We start by taking in space $Oxyz$ a domain Ω bounded by a closed surface S which is cut by any co-ordinate line in not more than two points (see Fig. 67).

Let us transform

$$I = \iiint_{\Omega} \frac{\partial Z}{\partial z}\, dx\, dy\, dz.$$

This is done by drawing the cylindrical surface orthogonally projecting domain Ω on to Oxy; it touches surface S in a curve L which divides it into two parts S_2 and S_1, each of which is cut by any straight line parallel to Oz in not more than one point. Let domain D be the projection of surfaces S_2 and S_1 (and domain Ω) on Oxy, whilst $z = z_2(x, y)$ and $z = z_1(x, y)$ are the equations of surfaces S_2 and S_1.

On integrating first with respect to z, then with respect to x and y over domain D, we get

$$I = \iiint_{\Omega} \frac{\partial Z}{\partial z}\, dx\, dy\, dz = \iint_{D} dx\, dy \int_{z_1(x,\,y)}^{z_2(x,\,y)} \frac{\partial Z}{\partial z}\, dz.$$

On carrying out the inner integration, we find that

$$I = \iint_{D} Z[x, y, z_2(x, y)]\, dx\, dy - \iint_{D} Z[x, y, z_1(x, y)]\, dx\, dy.$$

Since plane domain D is the projection of both S_2 and S_1 on Oxy, the double integrals on the right-hand side are the surface integrals of function $Z(x, y, z)$ over the upper sides of the surfaces.

Hence

$$I = \iint_{+S_2} Z(x, y, z)\, dx\, dy - \iint_{+S_1} Z(x, y, z)\, dx\, dy$$

$$= \iint_{+S_2} Z(x, y, z)\, dx\, dy + \iint_{-S_1} Z(x, y, z)\, dx\, dy,$$

i. e.

$$\iiint_{\Omega} \frac{\partial Z}{\partial z}\, dx\, dy\, dz = \iint_{S} Z\, dx\, dy, \tag{A}$$

the integration being over the *outside* of the entire surface S.

The formula still holds if the boundary of domain Ω — the surface S — happens to contain parts of the cylindrical surface with generators perpendicular to Oxy.

As in every previous case, we get rid of the condition imposed on surface S that it shall not be cut in more than two points by a co-ordinate line by dividing domain Ω into pieces and making use of the properties of triple and surface integrals. Formula (A) holds for domains having *any* continuous and measurable boundary S.

Similar proofs can be given of the equations

$$\iiint_{\Omega} \frac{\partial Y}{\partial y}\, dx\, dy\, dz = \iint_{S} Y\, dx\, dz\,, \tag{B}$$

$$\iiint_{\Omega} \frac{\partial X}{\partial x}\, dx\, dy\, dz = \iint_{S} X\, dy\, dz\,. \tag{C}$$

Addition of equations (A), (B), (C) term by term gives us *Ostrogradskii's formula* (*).

This proves the theorem.

Ostrogradskii's formula enables a triple integral, i.e. an integral over a spatial domain, to be replaced by a surface integral over the boundary of the domain, and conversely, a surface integral over a closed surface can be replaced by a triple integral over the domain bounded by the surface of integration.

In particular, let $X = x$, $Y = y$, $Z = z$. Then

$$\frac{\partial X}{\partial x} = \frac{\partial Y}{\partial y} = \frac{\partial Z}{\partial z} = 1$$

and we have

$$V = \iiint_{\Omega} dx\, dy\, dz = \frac{1}{3} \iint_{S} x\, dy\, dz + y\, dx\, dz + z\, dx\, dy\,,$$

i.e. we have obtained *an expression for a volume V in terms of a surface integral.*

II. We shall now solve the problem posed in Sec. 190 for surface integrals. *What conditions must be satisfied by functions X, Y, Z for the surface integral*

$$\iint_{S} X\, dy\, dz + Y\, dx\, dz + Z\, dx\, dy \tag{**}$$

to depend only on the boundary of the domain of integration, i.e. on the curve L bounding surface S — in other words, for it to remain the same whatever surface stretched over curve L we take as the domain of

integration? This requirement is equivalent to the following: integral (**) *over any closed surface must vanish.*

THEOREM. **The necessary and sufficient condition for integral (**) over a surface of integration stretched over a given curve and belonging to a singly-connected domain Ω to be independent of the surface is that functions $X(x, y, z)$, $Y(x, y, z)$, $Z(x, y, z)$, continuous together with their partial derivatives in domain Ω, satisfy at every point of the domain the equation**

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} = 0.$$

Proof. The sufficiency of the condition is clear from Ostrogradskii's formula; the necessity is proved by *reductio ad absurdum*, as in the similar cases of Secs. 185 and 190.

This is what we set out to prove.

CHAPTER XIV

DIFFERENTIAL EQUATIONS

1. Equations of the First Order

192. Equations with Separable Variables. The most effective and widespread method whereby mathematical analysis is used to solve concrete problems of pure and applied science is with the aid of differential equations.

The problems so far considered (see Sec. 115) have led to differential equations of an extremely simple form:

$$du = f(x)\,dx. \qquad (*)$$

The differential of one variable is expressed explicitly in terms of the other variable and its differential. Summation over all the "elements", i.e. integration of both sides:

$$\int_{u_0}^{u} du = \int_{x_0}^{x} f(x)\,dx\,,$$

yields an explicit expression for one variable (u) as a function of the other (x):

$$u = u_0 + \int_{x_0}^{x} f(x)\,dx \quad (= F(x))\,,$$

where $u_0 = F(x_0)$ is the value of u corresponding to the given value $x = x_0$.

If we find the connection between the differentials of two variables x and u in accordance with the conditions of the problem, we very often arrive at a differential equation of the form

$$f_1(u)\,du = f_2(x)\,dx, \qquad (**)$$

where $f_1(u)$ and $f_2(x)$ are known functions of their arguments. Differential equations of this type are described as having *separated variables*. This is because each variable only occurs on one side of

the equation, along with its differential. Equation (*) is a particular case of equation (**) ($f_1(u) \equiv 1$).

Definition. Differential equations which reduce to form (**) *by multiplication of both sides by the same expression are called* differential equations with separable variables.

This is the case, for instance, with

$$\frac{f_1(u)}{f_2(x)} = \frac{dx}{du};$$

the variables are not yet separated but can be made so by multiplying both sides by $f_2(x)\,du$, whence we arrive at equation (**). Since one differential expression ($f_1(u)\,du$) is identically equal to the other ($f_2(x)\,dx$), their integrals over the respective intervals of variation of u and x must also be equal:

$$\int_{u_0}^{u} f_1(u)\,du = \int_{x_0}^{x} f_2(x)\,dx.$$

To obtain a definite solution of the problem, we have to know in advance the so-called *initial condition*, i.e. a pair of corresponding numerical values of u and x (u_0 and x_0).

After carrying out the integrations we get a relationship between x and u which no longer contains their differentials:

$$F_1(u) - F_1(u_0) = F_2(x) - F_2(x_0) \quad (F_1' = f_1,\ F_2' = f_2).$$

This equation defines u as an implicit function of x.

It is often convenient to make use of indefinite integrals. Equation (**):

$$f_1(u)\,du = f_2(x)\,dx$$

gives us

$$\int f_1(u)\,du = \int f_2(x)\,dx.$$

On carrying out the integrations, we get a connection between variables x and u:

$$F_1(u) = F_2(x) + C$$

($F_1' = f_1$, $F_2' = f_2$, C is an arbitrary constant), defining u as an implicit function of x and depending on the arbitrary constant C. This function satisfies the equation (i.e. turns it into an identity) for any value of C.

We shall consider two problems from physics, the solutions of which are given by differential equations with separated variables.

1. THE IMPOVERISHMENT OF A SOLUTION. A vessel contains 100 l of solution containing 10 kg of pure salt. The solution is impoverished by fresh water flowing into the vessel at a uniform rate of 3 l/min, and by solution flowing out at a uniform rate of 2 l/min. We want to know how much pure salt remains in the solution after the process has continued for an hour.

We consider the process at some arbitrary instant t (minutes); let x kg of salt remain in the solution at this instant. Since the volume has been increased by $3t$ l and decreased by $2t$ l after t min, the volume will be $(100 + t)$ l, whilst the salt concentration is $x/(100 + t)$ (the solution is always kept homogeneous). If t receives the increment dt, x receives the increment Δx, expressing the amount of salt leaving the vessel in the time interval from t to $t + dt$.

We extract the principal part of this increment (dx) by supposing the process to be uniform in the infinitesimal time interval $[t, t + dt]$. If the process were uniform for unit time (1 min) $2x/(100 + t)$ kg of salt would escape (for 2 l solution leave in 1 min, and each litre contains $x/(100 + t)$ kg salt). But we assume uniformity for dt min, so that

$$\frac{2x}{100 + t}\,dt = -\,dx$$

(the minus sign is taken because $dx < 0$). Therefore

$$\frac{2\,dt}{100 + t} = -\frac{dx}{x}.$$

We have obtained a differential equation with separated variables, which is in fact the equation of the problem.

It is important to notice that one condition of the problem—the initial condition that there are 10 kg salt at $t = 0$—is not used in forming the differential eauation. The differential equation thus holds for any similar process with any initial quantity of salt, and not merely for the present process. The initial condition further informs us as to which of the possible solutions of the differential equation satisfies *all* the conditions of the problem.

We obviously have in our case:

$$\int_0^t \frac{2}{100 + t}\,dt = -\int_{10}^{x} \frac{dx}{x};$$

whence

$$2 \ln \frac{100 + t}{100} = \ln \frac{10}{x},$$

or

$$x = \frac{10 \cdot 100^2}{(100 + t)^2}.$$

This function is the solution of the differential equation with the given initial condition. We should have to change the lower limits of the integrals with different initial conditions.

If we take the indefinite integral of

$$\frac{2\,dt}{100 + t} = -\frac{dx}{x},$$

we get

$$2 \ln (100 + t) = \ln \frac{C}{x},$$

whence

$$x = \frac{C}{(100 + t)^2}.$$

This solution describes the process with any initial amount of salt. The given condition: $x = 10$ at $t = 0$ gives us a numerical value for C:

$$10 = \frac{C}{(100 + 0)^2}, \quad \text{i.e.} \quad C = 10 \cdot 100^2,$$

so that

$$x = \frac{100 \cdot 100^2}{(100 + t)^2}.$$

On setting $t = 60$ (min), we obtain the required result:

$$X = \frac{10 \cdot 100^2}{160^2} \approx 3{\cdot}9 \text{ kg}.$$

2. THE SHAPE OF A ROTATING LIQUID. A cyclindrical vessel containing a liquid rotates uniformly with angular velocity ω about the axis of the cylinder. It is required to find the shape of the free surface of the liquid.

We find the curve formed by the section of the free surface of the liquid by a plane through the axis of rotation. The curve will obviously be the same whatever plane we take through the axis.

We take the system of co-ordinates shown in Fig. 68. Two forces act at the point $M(x, z)$: the force of gravity

$$MP_1 = mg$$

and the centrifugal force

$$MP_2 = \frac{mv^2}{x} = m\omega^2 x.$$

Since the point is at rest with respect to the liquid as a whole, the resultant of these two forces must be balanced by the pressure on

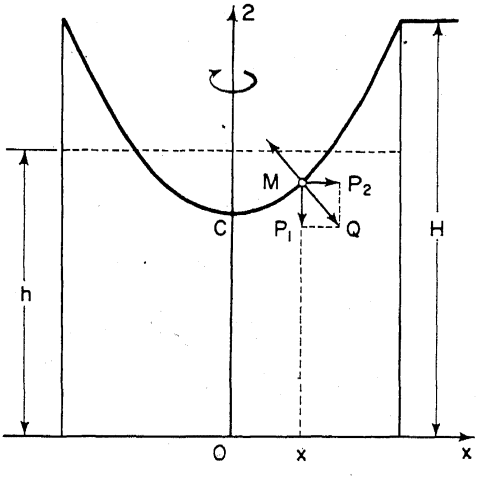

FIG. 68

the liquid, which acts perpendicularly to the free surface. It follows from this that the normal to the curve at M forms an angle α with OX such that

$$\tan\alpha = -\tan \angle P_2MQ$$
$$= -\frac{P_2Q}{P_2M} = -\frac{mg}{m\omega^2 x} = -\frac{g}{\omega^2 x}.$$

But on the other hand,

$$\tan\alpha = -\frac{1}{z'},$$

i.e.

$$\frac{1}{z'} = \frac{g}{\omega^2 x},$$

whence

$$x\,dx = \frac{g}{\omega^2}\,dz.$$

The problem has reduced to finding the equation of a curve from its differential property given by the last differential equation. We find on solving this equation:

$$z = \frac{\omega^2}{2g} x^2 + C,$$

i.e. we get a parabola symmetric with respect to Oz.

The free surface of the liquid thus has the form of a paraboloid of revolution.

Our answer is not quite definite—the equation contains an arbitrary constant Q (the height of the lowest point of the free surface above the bottom of the vessel). To find C, we need to be given a supplementary condition. This can be done in various ways. For instance, some point of the free surface can be determined experimentally, say the highest point (on the wall of the vessel). However, C may easily be found without experiment if we know the height h of the liquid in the cylinder at rest. The volume of liquid remains unchanged on rotation, so that

$$\pi r^2 h = \pi r^2 H - V,$$

where r is the radius of the cylinder, H the distance from the highest point of the free surface to the base, and V the volume of the body formed by rotation of our parabola about Oz. But

$$H = \frac{\omega^2}{2g} r^2 + C,$$

and

$$V = \pi \int_C^H x^2 dz = \frac{2g\pi}{\omega^2} \int_C^{\frac{\omega^2}{2g} r^2 + C} (z - C)\, dz = \frac{2g\pi}{\omega^2} \int_0^{\frac{\omega^2}{2g} r^2} u\, du = \frac{\pi\omega^2}{4g} r^4.$$

Consequently,

$$\pi r^2 h = \pi r^2 \left(\frac{\omega^2}{2g} r^2 + C \right) - \frac{\pi\omega^2}{4g} r^4 = \frac{\pi\omega^2}{4g} r^4 + \pi r^2 C,$$

i.e.

$$h = \frac{\omega^2}{4g} r^2 + C,$$

whence

$$C = h - \frac{\omega^2}{4g} r^2.$$

Thus, given the height of the liquid and radius of the cylinder we can find how far the level of the liquid will rise when the cylinder is rotated at constant angular velocity ω. If $h < \omega^2 r^2/4g$, part of the bottom will be free of liquid and the geometric vertex of the paraboloid will be below the bottom.

Certain technical devices are based on the above properties of rotating liquids.

193. General concepts. Differential equations are usually encountered in cases where it is impossible to establish directly a relationship between the variables themselves in the problem, whereas a relationship can be established between the differentials. A *differential law* is found for the process, from which the required relationship, i.e. the *integral law* of the process, follows by purely mathematical means. Some examples were considered in Sec. 192 which led to very simple relationships between the first order differentials (or derivatives), i.e. to differential equations with separable variables. However, the most elementary problems of physics, geometry and other sciences lead to various types of *differential equation*, in which differentials and derivatives of various orders occur.

Let us take, for instance, the simple problem of the fall of a body. When a body falls freely it encounters air resistance. We can represent this resistance as a force acting on the body in a direction opposite to that of the motion. As regards its magnitude, the two cases most commonly encountered are: *the resistance is proportional to the velocity; the resistance is proportional to the square of the velocity.* We have in the first case, in accordance with the fundamental equation of motion (see Sec. 59):

$$ms'' = mg - ks', \quad \text{or} \quad m\,d^2s = mg\,dt^2 - k\,ds\,dt,$$

and in the second

$$ms'' = mg - ks'^2, \quad \text{or} \quad m\,d^2s = mg\,dt^2 - k\,ds^2,$$

where m is the mass, g the acceleration due to gravity, k a coefficient of proportionality, and $s = s(t)$ the required distance of the body from the earth at the instant t.

Equations have been obtained in which the derivatives (or differentials) of the required function appear to the first and second orders. The function $s = s(t)$ describing the motion must in fact be found from these equations.

In view of what has been said, we shall set ourselves the task of considering methods of solving the more important types of differential equations to be found in applied mathematics. But, first of all, we need to discuss some general concepts.

Definition. **A *differential equation* is a relationship between the independent variable** (or variables), **the unknown function and its derivatives or differentials.**

The equation is an *ordinary* differential equation when it has one independent variable.

We shall only be concerned here with ordinary differential equations.

The order of a differential equation is the highest order of derivative (or differential) appearing in the equation.

Thus

$$x^2y'' + xy' + (x^2 - n^2)\,y = 0$$

is an equation of the second order; whilst

$$3y^2\,dy - 2x\,dx = 0$$

is of the first order.

Definition. **Any function satisfying a differential equation, i.e. that turns the equation into an identity when substituted in it, is called a *solution* of the equation.**

A solution of an ordinary differential equation has the form $y = y(x)$.

Definition. **An equation connecting the independent variable and the required function is called an *integral* of the differential equation.**

An integral of an ordinary differential equation has the form $u(x, y) = 0$.

For example, one of the solutions of the equation* $3y^2\,dy - 2x\,dx = 0$ is the function $y = x^{2/3}$, whilst one of the integrals is the equation

$$y^3 - x^2 = 0.$$

The integral of an equation defines the required function implicitly. The solution is found by solving the integral for the required function.

Every definite solution or integral of an equation has a corresponding graph called an integral curve *of the equation.*

The general form of first order equation is

$$f(x, y, y') = 0,$$

where f is a given function of three arguments: the independent variable x, the unknown function y and its derivative y'.

If the equation is solved for the derivative $y' = dy/dx$, it can be written as

$$\frac{dy}{dx} = y' = f(x, y), \quad \text{or} \quad dy = f(x, y)\,dx, \tag{*}$$

* We shall refer in future to a differential equation as simply an equation, provided it causes no confusion.

or in the more symmetric form:

$$X(x, y)\,dx + Y(x, y)\,dy = 0, \qquad (**)$$

where X and Y are known functions of arguments x and y.

Equation (**) will have separable variables if

$$X = \varphi_1(x)\,\psi_1(y) \quad \text{and} \quad Y = \varphi_2(x)\,\psi_2(y).$$

The following theorem holds for the general first order equation (*), solved for the derivative.

EXISTENCE AND UNIQUENESS THEOREM. **If function $f(x, y)$ is continuous in a domain containing the point (x_0, y_0), there exists a function $y = y(x)$ satisfying equation (*) and taking the value y_0 for $x = x_0$. If, in addition, the partial derivative $\partial f/\partial y$ is continuous, this solution of the equation is unique*.**

The condition that the required function $y(x)$ has the value y_0 for $x = x_0$ is known as the initial condition *of the equation;* it is written as: $y|_{x=x_0} = y_0$.

Geometrically, the theorem asserts the existence of a unique integral curve of the equation passing through point $P(x_0, y_0)$.

Definition. **The solution of the equation satisfying the given initial condition is called a *particular solution*, and the corresponding integral a *particular integral*.**

Suppose that the initial value $y = y_0$ corresponding to the given initial value $x = x_0$ is *arbitrary*. The solution of the equation will now contain an arbitrary constant $C = y_0$. But a solution can contain an arbitrary constant which is not the initial value of y. This leads us to the concept of *general solution*.

Definition. **The *general solution* of first order differential equation (*) is the solution $y(x, C)$ from which, given any possible** initial condition $y|_{x=x_0} = y_0$, a unique $C = C_0$ can be found such that**

$$y(x_0, C_0) = y_0.$$

The equation $u(x, y, C) = 0$ connecting the independent variable and the general solution, is termed the *general integral* of the differential equation.

* For the proof, see V.V. STEPANOV, *Course of differential equations* (*Kurs differentsial'nykh uravnenii*), 6th ed., Gost., 1953; also L.E. EL'SGOL'TS, *Ordinary differential equations* (*Obyknovennye differentsial'nye uravneniya*), p. 198, Gost., 1950.

** "Possible" initial condition means one with which the existence and uniqueness theorem holds for the solution.

The general solution (integral) is represented geometrically by a *one-parameter family of integral curves*, with not more than one curve of the family passing through each point of the Oxy plane.

Given the equation $u(x, y, C) = 0$ of a one-parameter family of curves, if we differentiate with respect to x:

$$\frac{\partial u(x, y, C)}{\partial x} = 0$$

and eliminate the arbitrary constant C from the two equations $u = 0$, $\partial u/\partial x = 0$, we obtain a first order *differential* equation $F(x, y, y') = 0$ for which the given *finite* equation is the general integral.

When $f(x, y) = f(x)$, i.e. when we have the elementary differential equation $dy = f(x)\,dx$, the existence theorem for the solution reduces to the existence theorem for the integral of a continuous function (Sec. 86), which in fact supplies the solution of the equation.

The function

$$y = \int_{x_0}^{x} f(x)\,dx + y_0$$

is a *particular* solution ($y|_{x=x_0} = y_0$), whilst

$$y = \int f(x)\,dx = \int_{x_0}^{x} f(x)\,dx + C$$

is the *general* solution. Given any initial condition $y|_{x=x_0} = y_0$, the choice of C is extremely simple: $C = y_0$.

We shall mention some simple examples of equations for which the existence and uniqueness theorem does not hold.

Example 1. We find the general solution of

$$y\,dx - x\,dy = 0, \quad \text{or} \quad y' = \frac{y}{x}.$$

On separating the variables and integrating, we get the general solution $y = Cx$, from which it is impossible to find a unique particular solution satisfying the initial condition $y|_{x=0} = 0$. An infinite set of such particular solutions exists. Any straight line through the origin is an integral curve of the equation. The uniqueness property does not hold, this being explained by the *discontinuity* of the right-hand side at $x = 0$.

Example 2. We find the general solution of

$$x\,dx + y\,dy = 0, \quad \text{or} \quad y' = -\frac{x}{y}.$$

Integration gives the general integral $x^2 + y^2 = C$, from which it is evident that the differential equation has in general no solutions satisfying the initial condition $y|_{x=0} = 0$. The family of integral curves is the set of concentric circles with centre at the origin. The absence of a particular solution is explained as above.

Seeking the solution (integral) of a differential equation is described as *integrating the equation.* This is generally (though not always) done by using some familiar operation of integration.

We shall regard a differential equation as having been integrated if: (1) *its solution has been found in a finite form,* or (2) *its solution is to be found by obtaining from a finite (not differential) equation a function given implicitly by the equation* (i.e. the integral of the equation has been obtained), or finally (3) *it only remains to take the integrals of known functions in order to obtain the solution, independently of whether or not these integrals are expressed with the aid of elementary functions.*

In the last case the integration of the differential equation is said to *reduce to a quadrature.*

The problem of forming the *differential equation* of a given one-parameter family of curves is the converse of that of integrating a differential equation; as indicated above, it is solved with the aid of *differentiation* and elimination of the arbitrary constants.

It must be noted that, when a concrete physical or geometrical problem is being investigated with the aid of a differential equation, a particular and not the general solution is required, such that all the conditions of the problem are satisfied.

194. Equations Reducible to Equations with Separable Variablese We consider the simplest types of first order equation, reducibly to equations with separable variables and therefore integrable by quadratures.

I. HOMOGENEOUS EQUATIONS.

Definition. **The equation**

$$y' = f(x, y)$$

is said to be *homogeneous* if $f(x, y)$ can be written as a function of the ratio of its arguments:

$$f(x, y) = \varphi\left(\frac{y}{x}\right).$$

For example, the equation

$$(xy - y^2)\,dx - (x^2 - 2xy)\,dy = 0$$

is homogeneous, since

$$f(x, y) = \frac{xy - y^2}{x^2 - 2xy} = \frac{\dfrac{y}{x} - \left(\dfrac{y}{x}\right)^2}{1 - 2\dfrac{y}{x}}.$$

The variables in a homogenous equation are not in general separable. But the homogeneous equation can be reduced to an equation with separable variables with the aid of a simple substitution for the unknown function y in terms of a new function u in accordance with the formula

$$\frac{y}{x} = u \quad \text{or} \quad y = xu.$$

We have, in fact:

$$y' = u + xu',$$

and the equation $y' = \varphi(y/x)$ becomes

$$u + xu' = \varphi(u), \quad \text{i.e.} \quad x\frac{du}{dx} = \varphi(u) - u.$$

Hence

$$\frac{du}{\varphi(u) - u} = \frac{dx}{x}$$

(the variables have been separated), which gives after integration:

$$\int \frac{du}{\varphi(u) - u} = \ln x.$$

This equation is in fact the integral. On finding from this the expression for u as a function of x and returning to variable $y = xu$, we obtain the solution of the homogeneous equation.

It is impossible in most cases to find an explicit expression for u. In this case we substitute y/x for u after integration of the left-hand side and arrive at the *integral* of the equation.

We are assuming, of course, that $\varphi(u) - u \not\equiv 0$. If $\varphi(u) \equiv u$, then $\varphi(y/x) \equiv y/x$, and no transformations are required, since the original equation has separable variables.

There is no need to memorize the formulae obtained above; it is easy to do the full working in each individual case.

Example. Find the integral of the homogeneous equation

$$y' = \frac{xy - y^2}{x^2 - 2xy}.$$

The substitution $y = xu$ leads to the equation

$$u + xu' = \frac{u - u^2}{1 - 2u},$$

or

$$\frac{du}{dx} = \frac{1}{x}\left(\frac{u - u^2}{1 - 2u} - u\right) = \frac{1}{x} \cdot \frac{u^2}{1 - 2u}.$$

We find on separating the variables:

$$\frac{1 - 2u}{u^2}\, du = \frac{dx}{x},$$

whence

$$\frac{1}{u} + 2\ln u = \ln \frac{C}{x}, \quad \text{or} \quad \ln\left(e^{\frac{1}{u}} u^2\right) = \ln \frac{C}{x}$$

i.e.

$$e^{\frac{1}{u}} u^2 = \frac{C}{x}.$$

On returning to variable y, we arrive at the general integral

$$\frac{y^2}{x}\, e^{\frac{x}{y}} = C.$$

This is actually the general integral of the equation, since a single value of C corresponds to any given initial condition $y|_{x=x_0} = y_0$, apart from the case when $x_0 = 0$:

$$C_0 = \frac{y_0^2}{x_0}\, e^{\frac{x_0}{y_0}}.$$

A large number of geometrical and physical problems are solved with the aid of homogeneous equations.

II. LINEAR EQUATIONS. A second commonly encountered type of first order equation is the linear equation.

Definition. **An equation of the form**

$$y' + p(x)\, y = q(x), \qquad (*)$$

i.e. linear in the required function and its derivative, is described as *linear*.

Here $p(x)$ and $q(x)$ are known functions of the independent variable.

Equation (*) reduces to an equation with separable variables by means of the following expedient. We write y as the product of two functions: $y = uv$. Naturally, one of them can be chosen in an entirely arbitrary manner; the second then has to be defined as a function of the first such that their product satisfies the given linear equation. We use the free choice of one of functions u or v so as to simplify as far as possible the equation obtained after the substitution.

We find the derivative y' from the equation $y = uv$:

$$y' = u'v + v'u.$$

On substituting in equation (*) we get

$$u'v + v'u + puv = q, \quad \text{or} \quad u'v + u(v' + pv) = q.$$

We take as v *any* particular solution of the equation

$$v' + pv = 0. \tag{**}$$

We now obtain the equation for u:

$$u'v = q. \tag{***}$$

Thus equation (*) is replaced by the two equations (**) and (***), each of which has separable variables.

We first find v from equation (**). We have on separating the variables:

$$\frac{dv}{v} = -p\,dx,$$

whence

$$\ln v = -\int p\,dx \quad \text{and} \quad v = e^{-\int p\,dx}.$$

The indefinite integral is taken here to mean *any* primitive of function $p(x)$.

Knowing v, we can now find u from equation (***):

$$\frac{du}{dx} = \frac{q}{v} = qe^{\int p\,dx}, \quad du = qe^{\int p\,dx}dx,$$

i.e.

$$u = \int qe^{\int p\,dx}dx.$$

The required function y is obtained from u and v:

$$y = uv = e^{-\int p\,dx}\int qe^{\int p\,dx}dx.$$

The integral on the line contains an arbitrary constant. The position is not affected if the integral in the exponent is also regarded as a set of primitives. The arbitrary constant supplied by this latter integral cancels out since one factor contains it in the denominator and one in the numerator.

The above formula gives the general solution of linear equation (*).

The problem can be solved with the aid of definite integrals with variable upper limits. In this case:

$$v = e^{-\int_{x_0}^{x} p\,dx}, \quad u = \int_{x_0}^{x} q e^{\int_{x_0}^{x} p\,dx} + C,$$

$$y = e^{-\int_{x_0}^{x} p\,dx}\left(\int_{x_0}^{x} q e^{\int_{x_0}^{x} p\,dx}\,dx + C\right).$$

The particular solution corresponding to the initial condition $y|_{x=x_0} = y_0$ is now obtained with $C = y_0$.

As before, there is no need to remember the general formula. We take note of the method and apply it in each case that comes along.

Example. $y' + y/x = \sin x/x$.

We put $y = uv$; then $y' = u'v + v'u$. We have:

$$u'v + v'u + \frac{1}{x}uv = \frac{\sin x}{x}, \quad \text{or} \quad u'v + u\left(v' + \frac{v}{x}\right) = \frac{\sin x}{x}.$$

Let $v' + \frac{v}{x} = 0$. Then $\frac{dv}{v} = -\frac{dx}{x}$, i.e. $v = \frac{1}{x}$. Therefore $\frac{u'}{x} = \frac{\sin x}{x}$, whence $u' = \sin x$, i.e. $u = -\cos x + C$. We finally get

$$y = uv = \frac{1}{x}\left(-\cos x + C\right).$$

Needless to say, we get the same expression if the general formula is used.

Linear equations are often encountered in practice. For instance, the flow of alternating current in a circuit containing inductance is described by a linear equation. Let $v = v(t)$ be the alternating voltage, $i = i(t)$ the current, R the circuit resistance and L the inductance (R and L are constants). We know from physics that the total voltage is the sum of the voltage across

the inductance, equal to $L di/dt$, and that across the resistance, equal to Ri. Thus

$$L\frac{di}{dt} + Ri = v.$$

Knowing $v(t)$, R and L, and the initial current $i|_{t=0} = i_0$, we can find from this linear equation the current as a function of time. We leave it to the reader to show that, if the voltage is sinusoidal: $v = A\sin\omega t$, the current is the sum of a sinusoidal part of the same frequency ω and a "damped" part which is neglected in practice.

III. EQUATIONS REDUCIBLE TO HOMOGENEOUS AND LINEAR TYPES. General methods can only be given in individual cases for reducing a differential equation to one with separable variables.

The first order homogeneous and linear equations considered in I and II are examples of such cases.

It sometimes happens that an equation is not of a familiar type but can be reduced to one for which the method of solution is known with the aid of a special device (substitution of the variables). It is impossible to give rules for finding suitable sustitutions in all possible cases. As in the integration of functions, the skill acquired by solving a large number of problems is of great value in finding suitable substitutions.

We shall consider some examples of first order equations reducible to homogeneous and linear equations.

Example 1.

$$y' = f\left(\frac{ax + by + c}{a_1x + b_1y + c_1}\right).$$

We put

$$x = x_1 + \alpha, \quad y = y_1 + \beta$$

with as yet undetermined α and β. Since $dy/dx = dy_1/dx_1$, we get

$$y_1 = f\left(\frac{ax_1 + by_1 + a\alpha + b\beta + c}{a_1x_1 + b_1y_1 + a_1\alpha + b_1\beta + c_1}\right).$$

We choose α and β so that

$$a\alpha + b\beta + c = 0 \quad \text{and} \quad a_1\alpha + b_1\beta + c_1 = 0.$$

This system is compatible provided $a/a_1 \neq b/b_1$ (or $ab_1 - ba_1 \neq 0$). Now

$$y_1' = f\left(\frac{ax_1 + by_1}{a_1x_1 + b_1y_1}\right) = f\left(\frac{a + b\dfrac{y_1}{x_1}}{a_1 + b_1\dfrac{y_1}{x_1}}\right) = \varphi\left(\frac{y_1}{x_1}\right),$$

and we arrive at a homogeneous equation. The required solution is obtained by solving it and returning to the original variables.

If $a/a_1 = b/b_1\ (= k)$, the variables can be separated at once in the given equation by substituting $y_1 = a_1x + b_1y$. For,

$$y' = \frac{1}{b_1}y_1' - \frac{a_1}{b_1} = f\left(\frac{ky_1 + c}{y_1 + c_1}\right) = \varphi(y_1),$$

whence

$$\frac{dy_1}{a_1 + b_1\varphi(y_1)} = dx.$$

Example 2. $$y' + py = qy^m.$$

This is known as *Bernoulli's type of equation*. It differs from a linear equation in having a power m $(m \neq 1,\ m \neq 0)$ of the required function y on the right-hand side.

We can reduce Bernoulli's to a linear equation by simple transformations.

We divide both sides by y^m:

$$y'y^{-m} + py^{1-m} = q.$$

We substitute $y^{1-m} = u$; then $u' = (1 - m)y^{-m}y'$, and the equation becomes

$$\frac{u'}{1 - m} + pu = q.$$

Having integrated this linear equation, we find the solution of the Bernoulli equation from the formula $y = u^{1/1-m}$.

195. Exact Differential Equations. The Integrating Factor.

I. EXACT DIFFERENTIAL EQUATIONS. We return to the first order equation which is solved for the derivative of the unknown function and has the general form

$$X(x, y)\,dx + Y(x, y)\,dy = 0. \qquad (*)$$

If the differential expression on the left-hand side is the total differential of some function $u(x, y)$, equation (*) *is called an exact differential equation.* As we know from Sec. 186, $X\,dx + Y\,dy$ is a total differential if $\partial X/\partial y = \partial Y/\partial x$.

Equation (*) can in this case be written as

$$du(x, y) = 0,$$

and its general integral is

$$u(x, y) = C,$$

where C is an arbitrary constant. Hence integration of equation (*) amounts to seeking the "primitive" of the left-hand side. If we make use of one of the expressions for this "primitive" obtained in Sec. 187, we get the general integral of (*) as

$$\int_{x_0}^{x} X(x, y)\,dx + \int_{y_0}^{y} Y(x_0, y)\,dy = C. \qquad (**)$$

We shall now deduce the formula for the integral of an exact differential equation by a different method, which does not depend on the concept of line integral. The method serves at the same time as a new* proof of the sufficiency (and necessity) of the condition $\partial X/\partial y = \partial Y/\partial x$ for $X\,dx + Y\,dy$ to be a total differential.

We seek the function $u(x, y)$, the total differential of which is equal to $X\,dx + Y\,dy$, or what amounts to the same thing, satisfies the equations

$$\frac{\partial u}{\partial x} = X(x, y), \quad \frac{\partial u}{\partial y} = Y(x, y). \qquad (A)$$

Integration of the first equation gives

$$u(x, y) = \int_{x_0}^{x} X(x, y)\,dx + \varphi(y), \qquad (B)$$

where x_0 is an arbitrary value of x and $\varphi(y)$ an arbitrary function of y. Our problem amounts to choosing $\varphi(y)$ so that the $u(x, y)$ given by (B) now satisfies the second of equations (A): $\partial u/\partial y = Y$. We show that such a function $u(x, y)$ exists provided $\partial X/\partial y = \partial Y/\partial x$, and we find $u(x, y)$.

We differentiate equation (B) with respect to y. We get by Leibniz's rule (see Sec. 181):

$$\frac{\partial u}{\partial y} = \int_{x_0}^{x} \frac{\partial X}{\partial y}\,dx + \varphi'(y),$$

or, since $\partial u/\partial y = Y$, $\partial X/\partial y = \partial Y/\partial x$,

$$Y(x, y) = \int_{x_0}^{x} \frac{\partial Y}{\partial x}\,dx + \varphi'(y) = Y(x, y) - Y(x_0, y) + \varphi'(y),$$

whence

$$\varphi'(y) = Y(x_0, y).$$

* The first proof was based on Green's formula (see Sec. 186).

We see that $\varphi'(y)$ depends only on y. Integration gives:

$$\varphi(y) = \int_{y_0}^{y} Y(x_0, y)\, dy,$$

where y_0 is an arbitrary value of y.

On substituting the expression got for $\varphi(y)$ in equation (B) and equating $u(x, y)$ to an arbitrary constant, we arrive at formula (**).

If $\partial X/\partial y \neq \partial Y/\partial x$, $Y - \int_{x_0}^{x} (\partial X/\partial y)\, dx$ will depend on x, which contradicts the independence of $\varphi'(y)$ on x. Hence there exists no function $u(x, y)$ satisfying both equations (A), and $X\,dx + Y\,dy$ is not a total differential.

Example. We find the general integral of the equation

$$(2x + y)\, dx + (x - 4y)\, dy = 0.$$

Since

$$\frac{\partial}{\partial y}(2x + y) = \frac{\partial}{\partial x}(x - 4y) = 1,$$

the left-hand side is the total differential of some function $u(x, y)$. We have:

$$\frac{\partial u}{\partial x} = 2x + y,$$

whence

$$u(x, y) = \int_{0}^{x} (2x + y)\, dx + \varphi(y) = x^2 + yx + \varphi(y).$$

Furthermore,

$$\frac{\partial u}{\partial y} = x + \varphi'(y) = x - 4y$$

i.e. $$\varphi'(y) = -4y, \quad \text{i.e.} \quad \varphi(y) = -2y^2.$$

Consequently,

$$u(x, y) = x^2 + xy - 2y^2 = C.$$

The differential equation solved here is homogeneous, so that it can also be integrated by the method described in Sec. 194. We recommend the reader to compare the two methods.

II. INTEGRATING FACTOR. We turn to the integration of equation (*) when $X'_y \neq Y'_x$, i.e. when it is not an exact differential equation.

There always exists a function $M = M(x, y)$ such that, when both sides of equation (*) are multiplied by it, the equation becomes

an exact differential equation. The function with this property is called the integrating factor of equation (*).

On multiplying both sides of (*) by a function $M(x, y)$—which does not change the general solution of the equation—we get: $MX\,dx + MY\,dy = 0$. The necessary and sufficient condition for this last equation to be an exact differential equation is that

$$\frac{\partial(MX)}{\partial y} = \frac{\partial(MY)}{\partial x},$$

i.e.
$$M\frac{\partial X}{\partial y} + X\frac{\partial M}{\partial y} = M\frac{\partial Y}{\partial x} + Y\frac{\partial M}{\partial x},$$

or
$$Y\frac{\partial M}{\partial x} - X\frac{\partial M}{\partial y} = M\left(\frac{\partial X}{\partial y} - \frac{\partial Y}{\partial x}\right),$$

or finally, on dividing both sides by M,

$$Y\frac{\partial \ln M}{\partial x} - X\frac{\partial \ln M}{\partial y} = \frac{\partial X}{\partial y} - \frac{\partial Y}{\partial x}. \qquad (A_1)$$

Obviously, every function $M(x, y)$ satisfying equation (A_1) can be taken as an integrating factor of equation (*); (A_1) *is the differential equation of the integrating factor of equation* (*). There is a theorem which says that (A_1) has an infinite set of solutions, so that equation (*) in fact always has an integrating factor†. But we have made no practical progress as regards integrating (*), since solution of equation (A_1) is no easier than our original problem. At the same time a particular feature of equation (A_1) occasionally makes it possible to choose at any rate one function $M(x, y)$ satisfying it (and there is no need to find more than one), whence the search for the integral of equation (*) reduces to a quadrature.

Example. Find the general integral of (see Sec. 186)

$$(3xy - 2y^2 + 4y)\,dx + (2x^2 - 3xy + 4x)\,dy = 0.$$

This is not an exact differential equation since $\partial X/\partial y \neq \partial Y/\partial x$ here. We write down the equation for the integrating factors:

$$x(2x - 3y + 4)\frac{\partial \ln M}{\partial x} - y(3x - 2y + 4)\frac{\partial \ln M}{\partial y} = -(x + y).$$

† It follows from this that there was no element of chance about the transformation of sec. 188 (when solving the thermodynamic problem) of the expression $(c_p/R)p\,dv + (c_v/R)v\,dp$ to an exact differential by multiplying by the function $1/T = R/pv$. The reader may verify from equation (A_1) that $1/T$ is in fact an integrating factor of the equation $(c_p/R)p\,dv + (c_v/R)v\,dp = 0$.

It is easy to choose a function M satisfying this equation, say $M = xy$. Consequently xy is an integrating factor, and

$$xy(3xy - 2y^2 + 4y)\,dx + xy(2x^2 - 3xy + 4x)\,dy = 0$$

is an exact differential equation. Solving this gives us the general integral of the original equation:

$$x^3y^2 - x^2y^3 + 2x^2y^2 = C.$$

One type of equation for which it is easy to find an integrating factor is that which admits of an integrating factor M depending on only one of the variables. Let M be a function satisfying equation (A_1) and not depending on y. Now $\partial \ln M/\partial y = 0$, and we have the ordinary equation for M:

$$\frac{d \ln M}{dx} = \frac{\dfrac{\partial X}{\partial y} - \dfrac{\partial Y}{\partial x}}{Y}, \qquad (A_2)$$

which gives us $\ln M$, and therefore M, by means of a single quadrature. It is clear that the right-hand side of equation (A_2) must not depend on y. Conversely, if the right-hand side of (A_2) is independent of y, an integrating factor M exists which is independent of y and satisfies equation (A_2).

The situation is similar when an integrating factor exists which is independent of x. The necessary and sufficient condition for this is that

$$\frac{\dfrac{\partial Y}{\partial x} - \dfrac{\partial X}{\partial y}}{X}$$

be independent of x; in this case,

$$\frac{d \ln M}{dy} = \frac{\dfrac{\partial Y}{\partial x} - \dfrac{\partial X}{\partial y}}{X}.$$

Example 1. We take the equation

$$(2 + 2x - y^2)\,dx - 2y\,dy = 0.$$

Here $\partial X/\partial y - \partial Y/\partial x$ $(= -2y - 0 = -2y)$ is not zero, i.e. we are not concerned with an exact differential equation. But

$$\frac{\dfrac{\partial X}{\partial y} - \dfrac{\partial Y}{\partial x}}{Y} \left(= \frac{-2y}{-2y} = 1\right)$$

does not depend on y, so that an integrating factor can be found from equation (A_2), which reads in the present case:

$$\frac{d \ln M}{dx} = 1.$$

Hence

$$\ln M = x, \quad \text{i.e.} \quad M = e^x.$$

On multiplying by e^x, we arrive at the exact differential equation

$$e^x(2 + 2x - y^2)\,dx - e^x 2y\,dy = 0.$$

There is no difficulty in solving this, and we get

$$2xe^x - y^2e^x = C.$$

Example 2. In the case of

$$xy\,dx + (x^2 - y^2 + 1)\,dy = 0$$

$\partial X/\partial y - \partial Y/\partial x\,(= x - 2x = -x)$ is again non-zero. We observe that

$$\frac{\dfrac{\partial Y}{\partial x} - \dfrac{\partial X}{\partial y}}{X}\left(= \frac{x}{xy} = \frac{1}{y}\right)$$

depends only on y. Hence an integrating factor exists which is independent of x. The equation

$$\frac{d \ln M}{dy} = \frac{1}{y}$$

gives us $M = y$.

The equation

$$xy^2\,dx + y(x^2 - y^2 + 1)\,dy = 0$$

is an exact differential equation and we obtain for the general integral:

$$y^2\left(x^2 - \frac{y^2}{2} + 1\right) = C$$

2. Equations of the First Order (Continued)

196. Tangent Field. Approximate Solutions. If no success can be had from any of the special methods of solving the first order equation

$$y' = f(x, y) \qquad (*)$$

or from the general integrating factor method, or if the resulting working is too unwieldy, recourse can be had to approximate methods. We shall describe here *Euler's graphical method* and the *numerical integration method* that follows from it, together with *Chaplygin's method* and the *method of integration with the aid of series*. As a preliminary, we shall consider the geometrical meaning of first order equation (*).

Equation (*) defines at every point $P(x, y)$ of the domain of Oxy in which the existence and uniqueness theorem holds (see Sec. 193) the slope of the tangent (y') to the integral curve through the point $P(x, y)$. The slope can be represented graphically by a straight arrow from the point P having a slope equal to $f(x, y)$, the length of the arrow being of no significance. Equation (*) is said to establish in this way a *tangent field* in the Oxy plane.

The locus of points associated with the same slope ($y' = \text{const}$) is said to be an *isocline* of the equation (line of equal slopes). Obviously, we get the equation of an isocline corresponding to a given value $y' = C$ if this value is substituted in the differential equation:

$$C = f(x, y).$$

If C is arbitrary but constant, this is the equation of the *family of isoclines* of differential equation (*).

At every point of a given isocline—corresponding to one value of C—the tangents to the integral curves have the same direction.

It is evident that the problem of integrating a differential equation can be interpreted geometrically as: *find the curve satisfying the condition that the tangents to it have the same direction as the tangents of the field at the points of contact.*

We now turn to the approximate integration methods; we assume throughout that the existence and uniqueness theorem is satisfied.

I. EULER'S GRAPHICAL METHOD. By starting from the above geometrical interpretation, we can find approximately by a graphical method a particular solution of a differential equation corresponding to a given initial condition $y|_{x=x_0} = y_0$ and interval $[x_0, x]$.

The problem amounts to constructing an integral curve through the initial point $M_0(x_0, y_0)$.

This can be done approximately, without a preliminary plotting of the tangent field, by means of simple constructions exactly similar to those used for "graphical integration" of functions, i.e. solution of equation (*) in the particular case when $f(x, y) = f(x)$. We divide interval $[x_0, x]$ into a number n of sub-intervals by points $x_0, x_1, x_2, \ldots, x_{n-1}, x_n = x$ (Fig. 69). We draw straight lines parallel to Oy through the points of sub-division and carry out the following sequence of operations.

We work out the value of $f(x, y)$ at the point $M_0(x_0, y_0)$; $F(x_0, y_0)$ measures in accordance with equation (*) the slope of the curve at M_0. We construct the tangent from M_0 by taking a pole P of the graph on Ox to the left of the origin such that $OP = 1$ (the scale for OP need not be the same as the scale on the co-ordinate axes); we mark off along Oy the segment ON_0, equal to the number $f(x_0, y_0)$ on the OP scale, and draw the straight line PN_0. The direction of PN_0 will evidently be the required direction (tangent) of the curve at M_0. We draw a straight line from M_0 parallel to PN_0 as far as its intersection with $x = x_1$. This gives us a point M_1, which we take as

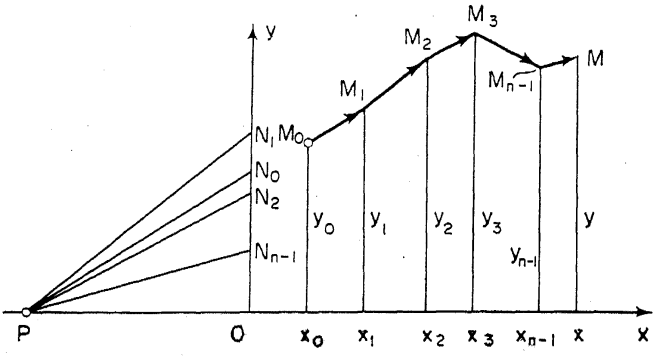

FIG. 69

the point of the integral curve corresponding to $x = x_1$. This construction implies replacing the arc of the curve in the sub-interval $[x_0, x_1]$ by a segment of its tangent at the initial point. Further, we work out $f(x, y)$ at the new point $M_1(x_1, y_1)$; $f(x_1, y_1)$ measures the slope of the curve at point M_1. We mark off on Oy a segment ON_1, equal to the number $f(x_1, y_1)$ on the OP scale, and draw the straight line joining P and N_1. Next we draw from M_1 a straight line parallel to PN_1 as far as its intersection with $x = x_2$. This gives us the point M_2, which we take as the point of the integral curve corresponding to $x = x_2$.

By proceeding in this way we find in turn the points of the curve corresponding to the points of sub-division $x_3, x_4, \ldots$ of interval $[x_0, x]$, until we arrive at the final point $M(x, y)$. The resulting step line $M_0M_1M_2 \ldots M_{n-1}M$ *approximately represents the integral curve through point* $M_0(x_0, y_0)$.

II. NUMERICAL INTEGRATION. We can translate into analytic language Euler's method for approximate integration of differential equation (*).

Obviously, the first operation gives the following relationship between the co-ordinates of points M_0 and M_1:

$$y_1 - y_0 = f(x_0, y_0)\,(x_1 - x_0); \tag{1}$$

the second operation leads to the similar relationship

$$y_2 - y_1 = f(x_1, y_1)\,(x_2 - x_1) \tag{2}$$

and so on; finally, the n-th operation gives

$$y - y_{n-1} = f(x_{n-1}, y_{n-1})\,(x - x_{n-1}). \tag{n}$$

These n equations enable us to work out successively the values of the unknown function at the points of sub-division of interval $[x_0, x]$. For, we find y_1 from the first equation for a given x_0, y_0 and chosen x_1, from the second y_2 for known x_1, y_1 and chosen x_2, and so on, until we arrive at the required value y.

On adding all n equations term by term, we get for y:

$$y = y_0 + f(x_0, y_0)(x_1 - x_0) + f(x_1, y_1)(x_2 - x_1) + \cdots + \\ + f(x_{n-1}, y_{n-1})(x - x_{n-1}).$$

The smaller the greatest of the sub-intervals, i.e. the greater n and the closer x to x_0, in general the more accurate the result.

As n increases indefinitely the last formula evidently becomes in the limit

$$y = y_0 + \int_{M_0 M} f(x, y)\, dx, \qquad (**)$$

where the line integral is over the integral curve $M_0 M$. But since this latter is unknown, expression (**) for y, whilst strictly accurate, cannot be used directly in practice for evaluating y, except in the case when $f(x, y) = f(x)$, i.e. when the strict solution is given by

$$y = y_0 + \int_{x_0}^{x} f(x)\, dx.$$

Thus the practical evaluation of the approximate value of y can be carried out with the aid of recurrence equations (1) — (n). This is one of the *methods of numerical integration* of equation (*). We speak of "numerical" integration because it amounts to finding the *numerical* value of the particular solution for some given value of the independent variable, and not to finding the actual *function*.

Example. Consider the equation

$$y' = xy^2 + 1.$$

None of the methods of solution by quadratures is applicable here. Let us find graphically the approximate solution of the equation in the interval [0, 1] with the initial condition $y|_{x=0} = 0$ and evaluate y at $x = 1$.

We divide [0, 1] into four parts by points $x_0 = 0$; $x_1 = 0{\cdot}25$; $x_2 = 0{\cdot}5$; $x_3 = 0{\cdot}75$; $x_4 = 1$ (Fig. 70). Let y'_0, y'_1, y'_2, y'_3 denote the values of y' at points x_0, x_1, x_2, x_3 respectively. Since $x_0 = 0$, $y_0 = 0$, we have $y'_0 = 1$, i.e.

$$y_1 = 1 \cdot (x_1 - x_0) = 0{\cdot}25, \qquad M_1(0{\cdot}25;\ 0{\cdot}25).$$

Further,

$$y'_1 = 0{\cdot}25 \cdot 0{\cdot}25^2 + 1 = 1{\cdot}016.$$

Therefore

$$y_2 = 0{\cdot}25 + 1{\cdot}016 \cdot 0{\cdot}25 = 0{\cdot}504, \quad M_2(0{\cdot}5;\ 0{\cdot}504).$$

Next,

$$y'_2 = 0{\cdot}5 \cdot 0{\cdot}504^2 + 1 = 1{\cdot}127$$

and

$$y_3 = 0{\cdot}504 + 1{\cdot}127 \cdot 0{\cdot}25 = 0{\cdot}786, \quad M_3(0{\cdot}75;\ 0{\cdot}786).$$

Finally,

$$y'_3 = 0{\cdot}75 \cdot 0{\cdot}786^2 + 1 = 1{\cdot}463$$

and

$$y = y_4 = 0{\cdot}786 + 1{\cdot}463 \cdot 0{\cdot}25 = 1{\cdot}152.$$

Thus $y = 1{\cdot}152$ is the required approximate value at $x = 1$ of the particular solution corresponding to the initial condition $y|_{x=0} = 0$.

Figure 70 illustrates the integral curve through (0, 0) drawn in accordance with our working.

III. CHAPLYGIN'S METHOD. An original method has been proposed by S. A. CHAPLYGIN (*see* Introduction, Sec. 5) for the approximate integration

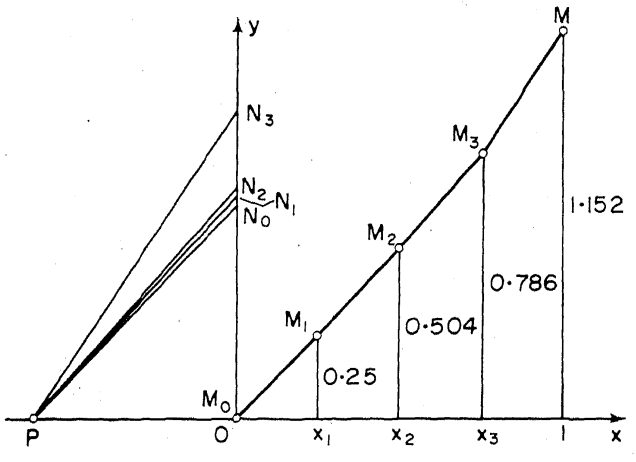

FIG. 70

of differential equations soluble with respect to the highest derivative. We shall sketch here the principles of the method as applied to first order equations.

The problem of finding an approximate solution $y = y(x)$ of equation (*):

$$y' = f(x, y)$$

with initial condition $y|_{x=x_0} = y_0$ will be solved in an interval $[x_0, x]$ if two functions $y = y_1(x)$, $y = y_2(x)$ can be indicated such that we have

$$y_1(x) \leqslant y(x) \leqslant y_2(x)$$

in the interval, where $y_1(x_0) = y_2(x_0) = y(x_0) = y_0$ and the difference $y_2(x) - y_1(x)$ does not exceed a specified error.

Thus "we must try to 'find the ribbon' ABC (Fig. 71) which is as narrow as possible and inside which the integral curve Ay must lie" (S. A. Chaplygin).

Chaplygin found a standard process for discovering functions $y_1(x)$ and $y_2(x)$ that came closer and closer to each other. This process is based on the following theorem of Chaplygin on "differential inequalities":

If the function $y = y_1(x)$ *is such that, in the interval* $[x_0, x]$,

$$y'_1(x) < f(x, y_1) \quad \text{and} \quad y_1(x_0) = y_0,$$

we have in this interval:

$$y_1(x) \leqslant y(x);$$

similarly, if $y_2'(x) > f(x, y_2)$, $y_2(x_0) = y_0$, *then* $y_2(x) \geqslant y(x)$.

N.E. ZHUKOVSKII (*see* Introduction, Sec. 5) pointed to the possibility of the simple geometrical proof given below for this theorem. Let $x > x_0$. Since $y_1(x_0) = y(x_0)$, we have $f[x_0, y_1(x_0)] = f[x_0, y(x_0)] = y'(x_0)$, so that $y_1'(x_0) < y'(x_0)$. This implies that the curve $y = y_1(x)$ has an inclination to Ox at the point $A(x_0, y_0)$ less than that of the curve $y = y(x)$. It follows from this that the curve $y = y_1(x)$ does not lie above the curve $y = y(x)$ in a neighbourhood measured from point A. If the curve $y = y_1(x)$ happens

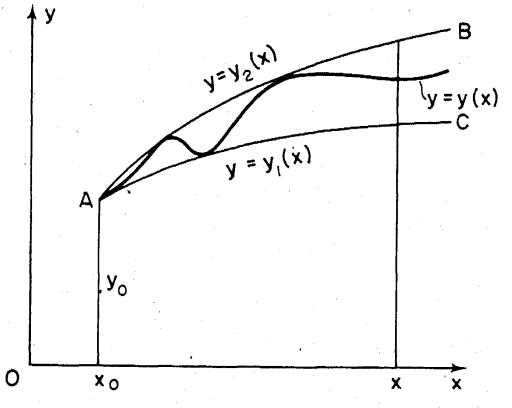

FIG. 71

to come above the curve $y = y(x)$ in some region of the interval $[x_0, x]$, the curves must intersect at certain points. Let A_1 denote the point of intersection that lies closest to A, and let x_1 be its abscissa. Then it is evident that $y = y_1(x)$ has an inclination at the point A_1 greater than that of $y = y(x)$, i.e. $y_1'(x_1) > y'(x_1)$, which contradicts our hypothesis, since $y_1'(x_1) < f[x_1, y_1(x_1)] = f[x_1, y(x_1)] = y'(x_1)$. Thus the inequality $y_1(x) \leqslant y(x)$ holds throughout $[x_0, x]$, in which the inequality $y_1'(x) < f(x, y_1)$ holds. A similar theorem holds for the case when $y_2'(x) > f(x, y_2)$.

Chaplygin's theorem is noteworthy in that it enables us to say from the sign of the expression $Y' - f(x, Y)$, where $Y(x)$ is a function subject only to the initial condition $Y(x_0) = y_0$, whether the integral curve of the equation $y' = f(x, y)$ lies above or below the curve $y = Y(x)$.

Suppose we know a pair of curves $y = y_1(x)$ and $y = y_2(x)$, $y_1(x_0) = y_2(x_0) = y_0$, "flanking" the curve $y = y(x)$; Chaplygin disovered a method of successively finding other pairs of curves with the same property and indefinitely approaching each other. No general rules can be laid down as regards finding the initial pair. But it is useful to recall the following when seeking such a pair: let functions $f_1(x, y)$ and $f_2(x, y)$ be such that

$$f_1(x, y) \leqslant f(x, y) \leqslant f_2(x, y),$$

and let the equations

$$y' = f_1(x, y), \quad y' = f_2(x, y)$$

have solutions

$$y_1(x) \quad \text{and} \quad y_2(x)$$

with initial conditions

$$y_1(x_0) = y_2(x_0) = y_0;$$

then

$$y_1(x) \leqslant y(x) \leqslant y_2(x).$$

For we have:

$$y_1' = f_1(x, y_1) \leqslant f(x, y_1) \quad \text{and} \quad y_2' = f_2(x, y_2) \geqslant f(x, y_2).$$

For further details, we recommend the reader to consult CHAPLYGIN'S *Collected Papers (Polnoe sobranie sochinnii)*, vol. III (1935), vol. I (1948), and special works on the approximate solution of differential equations*.

IV. INTEGRATION WITH THE AID OF SERIES. Power series can be applied for integrating equations in the same way as for integrating functions.

We assume that the solution is expressed as a power series and find the coefficients of the series from the given equation and the initial condition. If we can further prove that the series obtained is convergent, we can verify that it expresses the required solution (assuming that this is unique). There is often no need to prove this in the cases usually encountered.

A finite but large enough number of terms of the series gives us as close an approximation as may be desired to the solution, in the form of a Taylor polynomial.

We take the equation

$$y' = xy^2 + 1$$

with $y|_{x=0} = 0$.

We put

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots;$$

the initial condition gives us $a_0 = 0$. Substitution of the series in the equation gives

$$\begin{aligned} a_1 - 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1} + \cdots \\ = x(a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots)^2 + 1 \\ = x[a_1^2 x^2 + 2a_1 a_2 x^3 + (a_2^2 + 2a_1 a_3) x^4 + \\ + (2a_1 a_4 + 2a_2 a_3) x^5 + \cdots] + 1. \end{aligned}$$

On comparing coefficients of like powers, we get

$$a_1 = 1, \; a_2 = a_3 = 0, \; a_4 = \frac{1}{4}, \; a_5 = a_6 = 0, \; a_7 = \frac{1}{14}, \ldots$$

* See also L. E. EL'SGOL'TS, *Ordinary Differential Equations (Obyknovennye differentsial'nye uravneniya)*, p. 41, Gost., 1957.

The first terms of the series for the solution are therefore

$$y = x + \frac{x^4}{4} + \frac{x^7}{14} + \cdots$$

It should be noted that some earlier coefficients are usually more easily found from the equation itself as follows. We differentiate the equation several times:

$$y'' = y^2 + 2xyy',$$
$$y''' = 4yy' + 2xy'^2 + 2xyy'',$$
$$y^{IV} = 6y'^2 + 6yy'' + 6xy'y'' + 2xyy'''.$$

On setting $x = 0$ in all these equations and noting that $y(0) = 0$, we find successively: $y'(0) = 1, y''(0) = 0, y'''(0) = 0, y^{IV}(0) = 6, \ldots$

Hence

$$y(x) = x + \frac{6}{4!}x^4 + \cdots = x + \frac{x^4}{4} + \cdots$$

197. Singular Solutions. Clairaut's Equation.

I. SINGULAR SOLUTIONS. As we have seen (see Sec. 193), there are cases when there are several solutions, and not just one, corresponding to a given initial condition.

In other words, a point of the plane $P_0(x_0, y_0)$ can exist through which several integral curves pass, having a common tangent at their common point P_0. If the points of this sort form an integral curve, it is said to be *singular*.

Definition. **We describe as *singular* an integral curve of the first order differential equation**

$$f(x, y, y') = 0,$$

through each point of which there passes at least one further integral curve having the same tangent; the equation and function corresponding to the singular integral curve are termed the *singular integral* and *singular solution* of the equation.

A singular integral curve is not usually to be found in the general family of integral curves, i.e. *it is not a particular solution*—it cannot be obtained from the general solution for any value of the arbitrary constant. (Singular solutions are often defined as solutions not contained in the general solution of a differential equation).

If we have already found the general integral of a first order equation, it is easy to discover whether or not the equation has a

singular integral, and if so, what it is. All we need to know is whether or not the general family of integral curves of the equation has an envelope (see Sec. 164). The equation of the envelope is in fact the singular integral.

We shall prove a theorem in connection with this.

THEOREM. **If we know the general integral of the first order differential equation**

$$u(x, y, C) = 0,$$

elimination of the arbitrary constant C from this equation and the equation

$$\frac{\partial u(x, y, C)}{\partial C} = 0,$$

gives the singular integral*.

Proof. Elimination of C from the equation $u = 0$ and $\partial u/\partial C = 0$ in fact gives the equation of the envelope of the family of integral curves of the equation. This envelope is necessarily an *integral curve* and at the same time *singular*. For, by the property of the envelope, at every point of it the "element" (x, y, y') is the same as the "element" of the integral curve touching it at this point, so that the numbers x, y, y' turn the equation $f(x, y, y') = 0$ into an identity; consequently the function represented by the envelope satisfies the given differential equation. This solution is singular because a curve of the general family of integral curves passes through every point of it and has the same tangent.

This is what we wished to prove.

Example. Let us find the curve with constant normal equal to a. The problem leads us to the following differential equation:

$$|y| \sqrt{1 + y'^2} = a. \qquad (*)$$

The variables may be separated:

$$1 + y'^2 = \frac{a^2}{y^2}, \quad y' = \frac{1}{y}\sqrt{a^2 - y^2},$$

whence

$$\frac{y\,dy}{\sqrt{a^2 - y^2}} = dx.$$

We obtain on solving this equation:

$$-\sqrt{a^2 - y^2} = x + C.$$

* Assuming, of course, that elimination is possible and that the discriminant curve is in fact the envelope and not the locus of singular points.

The general integral is therefore

$$(x - C)^2 + y^2 = a^2,$$

i.e. geometrically, the family of circles of radius a with centres on Ox.

We equate the derivative with respect to C to zero:

$$-2(x - C) = 0.$$

Elimination of C gives the envelope: $y = \pm a$ (see Sec. 164). We have thus found a singular solution of equation (*). The function obtained in fact satisfies the equation, whilst it is not obtained from the general integral whatever the value of C. The straight lines $y = a$, $y = -a$ are singular integral curves (and they in fact satisfy the hypothesis since their normals are always equal to a), whilst they do not belong to the general family of integral curves.

The singular integral of $f(x, y, y') = 0$ can be found in certain cases without having a knowledge of the general integral. If the operations used for finding the general integral destroy the equivalence of the equations, we have to find by substitution which of the lost expressions for y satisfy the equation; these expressions may include *singular* solutions. Thus when finding the general integral of equation (*) we divide both sides by $\sqrt{a^2 - y^2}$ during the course of the working, which involves the loss of the solution $y = \pm a$. This latter is, in fact, the singular solution.

At the same time there are other cases where, in similar circumstances, the lost expression for y is not a singular solution. For instance, we divide both sides by y when solving the equation $y\,dx - x\,dy = 0$; but $y = 0$ is a particular, and not a singular solution.

II. CLAIRAUT'S EQUATION. An important class of first order differential equations not soluble for y' is typified by the *Clairaut equation**:

$$y = xy' + \psi(y').$$

The following device is used for integrating Clairaut's equation. We write p for y' for the sake of convenience:

$$y = xp + \psi(p)$$

* CLAIRAUT (1713–1765).

and differentiate the equation with respect to x:

$$p = p + xp' + \psi'(p)\, p',$$

whence

$$p'[x + \psi'(p)] = 0.$$

We get by equating each factor on the left to zero:

$$p' = 0, \quad x + \psi'(p) = 0.$$

The first equation gives $p = C$; substituting this in the origina equation gives the *general integral*

$$y = xC + \psi(C).$$

We see that the general integral of Clairaut's equation is a one parameter family of straight lines; it is got from the equation by replacing y' by an arbitrary constant C. Thus *the general family of integral curves of Clairaut's equation is the family of its isoclines* (see Sec. 196).

The second equation gives

$$x = -\psi'(p).$$

By combining this with the original equation

$$y = xp + \psi(p) = -p\psi'(p) + \psi(p),$$

we get the parametric form of an integral of the equation. Elimination of parameter p gives the *singular integral* of Clairaut's equation.

This is proved by verifying that the integral curve defined by it is the envelope of the general family of integral curves (i.e. of the straight lines $y = xC + \psi(C)$). We replace the letter p by C in the parametric equations (this does not affect the sense of the equations):

$$x + \psi'(C) = 0, \quad y = xC + \psi(C).$$

Elimination of C from these two equations leads to the integral in question. But one of them—the first—is got from the other—the second—by differentiation with respect to C, which in fact proves the assertion, since the second equation is actually the general integral.

The singular integral of Clairaut's equation is not a linear function, so that it is not contained in the general integral.

Clairaut's equation has no singular solution when $\psi'(p) = \text{const.}$, i.e. when $\psi(y')$ is a linear function of y'. In this case the equation has *separable variables*.

III. We are often more interested in the singular than the general integral when dealing with a concrete problem leading to a differential equation $f(x, y, y') = 0$. For instance, the general solution of Clairaut's equation corresponds to a family of straight lines, which is usually a trivial solution from the point of view of say geometrical problems; the important solution is given by the envelope of these straight lines.

Quite a number of geometrical problems lead to a Clairaut equation.

Example. Let a curve be such that the segment of every tangent to it cut off by the co-ordinate axes is equal to a constant a. Let us form the differential equation of the family of such curves. The equation of the tangent is

$$\eta - y = y'(\xi - x),$$

where ξ, η are the current co-ordinates of the tangent, x, y the co-ordinates of the point of contact; hence the segments cut off on the axes are

$$-\frac{y - y'x}{y'}, \quad y - y'x.$$

We thus have

$$\frac{(y - y'x)^2}{y'^2} + (y - y'x)^2 = a^2,$$

or

$$(y - y'x)\frac{\sqrt{1 + y'^2}}{y'} = a,$$

and we obtain a Clairaut equation for the family:

$$y = xy' + \frac{ay'}{\sqrt{1 + y'^2}}.$$

Its general solution is given by

$$y = Cx + \frac{aC}{\sqrt{1 + C^2}}.$$

This is the equation of the family of straight lines whose segments lying between the axes are equal to the given a. Such a solution is clearly trivial for the present problem. The solution of interest is the singular integral, i.e. geometrically, the envelope of the above family of straight lines; the reader may easily show that this is the *astroid* (see Sec. 83):

$$x^{2/3} + y^{2/3} = a^{2/3}.$$

IV. Clairaut's equation is a particular case of an equation of a more general type—*Lagrange's equation.*

Definition. A Lagrange equation is a first order differential equation linear in the independent variable and the function y:

$$y = x\varphi(y') + \psi(y').$$

Clairaut's equation is the particular case when $\varphi(y') \equiv y'$. Let $\varphi(y')$ be not identically equal to y'. We write p for y' and apply the same method as for Clairaut's equation. We have:

$$p = \varphi(p) + x\varphi'(p)\frac{dp}{dx} + \psi'(p)\frac{dp}{dx},$$

whence

$$p - \varphi(p) = [x\varphi'(p) + \psi'(p)]\frac{dp}{dx}. \qquad (*)$$

Since $p - \varphi(p)$ is not identically zero, we can rewrite (*) as

$$\frac{dx}{dp} + x\frac{\varphi'(p)}{\varphi(p) - p} = \frac{\psi'(p)}{p - \varphi(p)}.$$

This is a linear equation for x, regarded as a required function of p. On obtaining $x = \omega(p, C)$ by means of quadratures and substituting this in the original equation:

$$y = \omega(p, C)\varphi(p) + \psi(p),$$

we arrive at two parametric equations (with parameter p) for the general integral. Elimination of p from them gives the general integral as a direct relationship between y, x and C.

There may be a singular integral of the Lagrange equation, in addition to the general integral. Suppose that $\varphi(p) - p$ vanishes for some value $p = p_0$. Then $p = p_0$ satisfies equation (*). Substitution of this value of p in the initial equation gives the linear function

$$y = \varphi(p_0)x + \psi(p_0),$$

which is in fact a singular solution of the Lagrange equation.

198. Orthogonal and Isogonal Trajectories. Let

$$F_1(x, y, C) = 0.$$

be a one-parameter family of curves. We pose the problem of finding another family of curves $F_2(x, y, C) = 0$ such that any curve of the second family cuts a curve of the first at a constant angle. Problems of this type most often arise in mechanics and are known as *problems on isogonal trajectories.*

Definition. **Two families of curves $F_1(x, y, C) = 0$ and $F_2(x, y, C) = 0$ having the property that any two curves of the different families intersect at the same angle (α) are described as *families of isogonal trajectories* with respect to each other; in particular, when the angle**

of intersection is a right-angle ($\alpha = \frac{1}{2}\pi$), the families are of *orthogonal trajectories*.

We shall first of all consider families of orthogonal trajectories.

We can specify a family of curves by a differential equation instead of a finite equation $F_1(x, y, C) = 0$, the latter being in this case the general integral of the differential equation. The differential equation is got by eliminating parameter C from the given equation $F_1(x, y, C) = 0$ and the equation $\partial[F_1(x, y, C))]/\partial x = 0$, obtained from it by differentiation with respect to x (see Sec. 193). Thus let the differential equation of the family have the form $f_1(x, y, y') = 0$. The derivative y' gives the slope of the tangent to the curve at the point (x, y). Since the orthogonal curve passing through (x, y) intersects the first curve at a right-angle, the slope of the tangent to the orthogonal curve—written as y_1'—must be the reciprocal of y' taken with the opposite sign: $y' = -1/y_1'$. On substituting in the equation and omitting the subscript for y, which has no significance for what follows, we find the equation

$$f_1\left(x, y, -\frac{1}{y'}\right) = 0,$$

which can be written as

$$f_2(x, y, y') = 0.$$

This is the *differential equation of the family of orthogonal trajectories*. Its general integral $F_2(x, y, C) = 0$ gives the required finite equation of the family.

Example 1. Find the family of trajectories orthogonal to the straight lines $y = Cx$ through the origin.

The differential equation of this family of straight lines is evidently $y' = y/x$. Hence the differential equation of the orthogonal trajectories has the form

$$y' = -\frac{x}{y}.$$

We obtain on separating the variables and integrating:

$$x^2 + y^2 = C^2,$$

i.e. the equation of the family of concentric circles with centres at the origin. The obviousness of this result makes further explanations unnecessary.

Example 2. We take the family of confocal ellipses with foci at $(-1,0)$ and $(1,0)$. The equation of the family can be written as

$$\frac{x^2}{1+C}+\frac{y^2}{C}=1,$$

where $C>0$.

We find the differential equation. Differentiation with respect to x gives

$$\frac{x}{1+C}+\frac{yy'}{C}=0.$$

Having found C from this, we substitute in the original equation and arrive at the differential equation of the family of confocal ellipses:

$$(x+yy')(xy'-y)=y'.$$

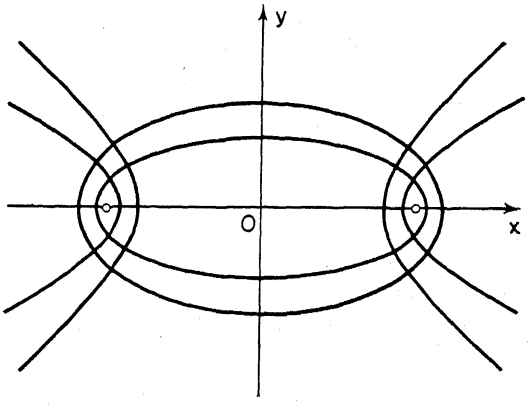

FIG. 72

We obtain the differential equation of the orthogonal trajectories if we replace y' here by $-1/y'$:

$$\left(x-\frac{y}{y'}\right)\left(-\frac{x}{y'}-y\right)=-\frac{1}{y'},$$

or

$$(x+yy')(xy'-y)=y',$$

i.e. the same equation. Thus its general integral will be

$$\frac{x^2}{1+C}+\frac{y^2}{C}=1.$$

The family of integral curves consists of confocal ellipses ($C>0$) and hyperbolas ($C<0$). We conclude from this that the required family of orthogonal trajectories is the family of confocal hyperbolas (with foci at the same points) (Fig. 72).

We turn to arbitrarily isogonal trajectories. If a curve of the second family is cut by a curve of the first at an angle α, their slopes y' and y'_1 must be related by

$$\frac{y'_1 - y'}{1 + y'y'_1} = \tan\alpha.$$

We can thus express y' in terms of y'_1 and $\tan\alpha$; on substituting this in the differential equation of the first family (and omitting the subscript of the derivative), we get the differential equation of the family of isogonal trajectories.

Example. Find the isogonal trajectories to the straight lines $y = Cx$. The differential equation of the family of straight lines is $y' = y/x$. On bringing this expression for the derivative into the relationship between the slopes then neglecting the subscript, we arrive at the differential equation of the isogonal trajectories:

$$\frac{y' - \dfrac{y}{x}}{1 + \dfrac{y}{x}y'} = \tan\alpha = k.$$

Hence

$$y' = \frac{y + kx}{x - ky}.$$

On solving this homogeneous equation in accordance with the general rule (Sec. 194, I), we get

$$\ln\sqrt{x^2 + y^2} = \frac{1}{k}\arctan\frac{y}{x} + \ln C.$$

We obtain on passing to polar co-ordinates ϱ, φ:

$$\varrho = Ce^{m\varphi},$$

where

$$m = \frac{1}{k} = \cot\alpha.$$

The isogonal trajectories are therefore logarithmic spirals, in other words, the only curves with the property that radius vectors from the origin cut them at a constant angle are the spirals $\varrho = Ce^{m\varphi}$. This property of logarithmic spirals was obtained directly in Sec. 56. We can now see that it completely characterizes the curves. The trajectories become orthogonal with $\alpha = \frac{1}{2}\pi$, and the family of logarithmic spirals degenerates ($m = 0$) to the family of concentric circles.

3. Equations of the Second and Higher Orders

199. General Concepts. We shall be chiefly concerned with differential equations of the second order, which have great importance in applied mathematics. The general concepts will be described, however, for equations of any order n.

We shall only encounter below n-th order differential equations containing the highest order derivative explicitly:

$$y^{(n)} = f(x, y, y', \ldots, y^{(n-1)}). \qquad (*)$$

The following theorem holds for these equations, as in the case $n = 1$.

EXISTENCE AND UNIQUENESS THEOREM*. **If the right-hand side of equation (*) — the function f — is continuous together with its partial derivatives with respect to arguments $y, y', \ldots, y^{(n-1)}$ in a domain containing the point $(x_0, y_0, y_0', \ldots, y^{(n-1)})$, the equation has a solution $y = y(x)$ which is unique and takes, along with its first $n-1$ derivatives, the given values $y(x_0) = y_0$, $y'(x_0) = y_0', \ldots, y^{(n-1)}(x_0) = y_0^{(n-1)}$ at $x = x_0$.**

The conditions indicating the values that must be taken by the required function y and its derivatives $y', y'', \ldots, y^{(n-1)}$ at the initial value $x = x_0$ are known as the initial conditions of the n-th order differential equation (or of the corresponding problem). They can be written briefly as

$$y|_{x=x_0} = y_0, \; y'|_{x=x_0} = y_0', \ldots, y^{(n-1)}|_{x=x_0} = y_0^{(n-1)}.$$

Definition. **A solution of an n-th order equation satisfying a given initial condition is called a *particular solution*, and the corresponding integral a *particular integral* of the equation.**

We shall now consider the initial values

$$y = y_0, \; y' = y_0', \ldots, \; y^{(n-1)} = y_0^{(n-1)},$$

corresponding to the initial value $x = x_0$ as *variables*. The solution will now obviously depend on these n variables. In the general case the solution can depend on n arbitrary parameters $C_0, C_1, \ldots, C_{n-1}$:

$$y = y(x, C_0, C_1, \ldots, C_{n-1}),$$

which leads us to the concept of the *general solution*.

Definition. **The *general solution* of n-th order differential equation (*) is the solution $y(x, C_0, C_1, \ldots, C_{n-1})$, from which, given any possible** initial conditions $y|_{x=x_0} = y_0, y'|_{x=x_0} = y_0', \ldots, y^{(n-1)}|_{x=x_0} = y_0^{(n-1)}$, unique values**

$$C_0 = C_{0,0}, \quad C_1 = C_{1,0} \ldots, \quad C_{n-1} = C_{n-1,0}$$

* See e.g. V. V. STEPANOV, *Course of Differential Equations* (*Kurs differentsial'nykh uravnenii*), 6th ed., Gost., 1953.

** i.e. conditions guaranteeing the existence and uniqueness of the solution.

can be found such that

$$y(x_0, C_{0,0}, \ldots, C_{n-1,0}) = y_0,\ y'(x_0, C_{0,0}, C_{1,0}, \ldots, C_{n-1,0}) = y'_0, \ldots, y^{(n-1)}(x_0, C_{0,0}, C_{1,0}, \ldots, C_{n-1,0}) = y_0^{(n-1)}.$$

The equation $u(x, y, C_0, C_1, \ldots, C_{n-1}) = 0$ connecting the independent variable and the general solution is called the *general integral* of the n-th order differential equation.

The geometric form of the general solution is an n-parameter family of integral curves.

Suppose we are given the equation $u(x, y, C_0, C_1, \ldots, C_{n-1}) = 0$ of an n-parameter family of curves; on differentiating this equation n times successively with respect to x and eliminating the arbitrary constants $C_0, C_1, \ldots, C_{n-1}$ from the $n + 1$ equations: $u = 0$, $\partial u/\partial x = 0$, $\partial^2 u/\partial x^2 = 0, \ldots, \partial^n u/\partial x^n = 0$, we get the n-th *order differential equation:* $F(x, y, y', \ldots, y^{(n)}) = 0$, for which the given *finite* equation is the general integral.

Whereas a first order equation expresses a property of the integral curves connected only with their directions (tangents), a second order equations expresses a property connected with their curvature as well as their directions.

Example. Find the curves for which the radius of curvature is constant a.

The condition of the problem leads at once to the second order differential equation

$$\frac{(1 + y'^2)^{3/2}}{y''} = a.$$

The symbol of the absolute magnitude is omitted from the expression for the radius of curvature, since it has no significance here. We integrate this equation by putting $y' = z$. Then $y'' = z'$, and we arrive at the first order equation with separable variables:

$$\frac{(1 + z^2)^{3/2}}{z'} = a \quad \text{or} \quad dx = a\frac{dz}{(1 + z^2)^{3/2}}.$$

We find that

$$x + C_1 = a\frac{z}{(1 + z^2)^{1/2}},$$

whence

$$z = y' = \frac{x + C_1}{\sqrt{a^2 - (x + C_1)^2}}.$$

We obtain on integrating again:

$$y + C_2 = -\sqrt{a^2 - (x + C_1)^2}, \quad \text{i.e.} \quad (x + C_1)^2 + (y + C_2)^2 = a^2.$$

We have found a relationship between x and y depending on two arbitrary constants C_1 and C_2 as the general integral of the given equation. This is the equation of the family of all circles of radius a. It follows from our solution that *the only curves with constant radius of curvature are circles.*

When solving problems of physics or geometry we usually want to find a particular, and not the general, solution of an n-th order differential equation, corresponding to all the conditions of the problem. It is most often obtained in practice, not from the general solution, but by finding the arbitrary constants of integration successively during the process of solving the equation (see Sec. 200, II).

200. Particular Cases. We shall consider the elementary types of higher order differential equations, reducible to first order equations and integrable by quadratures.

I. THE EQUATION OF THE FORM

$$y^{(n)} = f(x). \qquad (*)$$

We show that the solution of this equation is obtained by a single quadrature.

Since $y^{(n)} = (y^{(n-1)})'$, we have

$$y^{(n-1)} = \int_{x_0}^{x} f(x)\,dx + C_0,$$

where x_0 is any given value of x and C_0 is an arbitrary constant. We obtain on integrating again:

$$y^{(n-2)} = \int_{x_0}^{x} dx \int_{x_0}^{x} f(x)\,dx + C_0(x - x_0) + C_1.$$

On proceeding in this way, we eventually obtain

$$y = \underbrace{\int_{x_0}^{x} dx \int_{x_0}^{x} dx \ldots \int_{x_0}^{x}}_{n \text{ times}} f(x)\,dx + \frac{C_0}{(n-1)!}(x - x_0)^{n-1} +$$

$$+ \frac{C_1}{(n-2)!}(x - x_0)^{n-2} + \cdots + C_{n-2}(x - x_0) + C_{n-1}.$$

This is the general solution, given by an n-ple integral. For we can easily verify that the particular solution for initial conditions $y|_{x=x_0} = y_0$, $y'|_{x=x_0} = y_0'$, ..., $y^{(n-1)}|_{x=x_0} = y_0^{(n-1)}$ is obtained by assigning to the constant C the values $C_0 = y_0^{(n-1)}$, $C_1 = y_0^{(n-2)}$, ... $C_{n-2} = y_0'$, $C_{n-1} = y_0$. The integral term of the general solution

$$\int_{x_0}^{x} dx \int_{x_0}^{x} dx \ldots \int_{x_0}^{x} f(x)\, dx$$

is the particular solution which vanishes at $x = x_0$ along with its first $n - 1$ derivatives: $y(x_0) = y'(x_0) = \cdots = y^{(n-1)}(x_0) = 0$. But we know that this n-ple integral can be expressed as a single integral depending on the parameter x (see Sec. 180, Cauchy's formula):

$$\int_{x_0}^{x} dx \int_{x_0}^{x} dx \ldots \int_{x_0}^{x} f(x)\, dx = \frac{1}{(n-1)!} \int_{x_0}^{x} (x - z)^{n-1} f(z)\, dz.$$

Consequently, the general solution of equation (1) is given by a formula which contains only one quadrature:

$$y = \frac{1}{(n-1)!} \int_{x_0}^{x} (x - z)^{n-1} f(z)\, dz + \frac{C_0}{(n-1)!} (x - x_0)^{n-1} + \cdots$$
$$\cdots + C_{n-2}(x - x_0) + C_{n-1}.$$

The second order differential equation $y'' = f(x)$ is often encountered in dynamics. It gives the law of motion when the force can be expressed as a function of time only.

Example. Let the motion be along the Os axis under the action of a periodic force p, directed in opposition to the motion and depending on time in accordance with $p = -A\omega^2 \sin \omega t$, where $s|_{t=0} = 0$, $s'|_{t=0} = A\omega$. We find the equation of the motion, i.e. the position of the point s as a function of time t.

We have by the fundamental equation of mechanics:

$$s'' = -A\omega^2 \sin \omega t$$

(we take the mass m as unity for simplicity).

We write down the solution (it is more convenient not to pass to single integrals here):

$$s = \int_0^t dt \int_0^t (-A\omega^2 \sin \omega t)\, dt + A\omega t = A \sin \omega t.$$

Thus the motion is a *harmonic vibration* of the same frequency as the oscillating force. The differential equation of the motion can be written as $s'' = -\omega^2 s$.

II. THE SECOND ORDER EQUATION OF THE FORM

$$y'' = f(x, y'). \qquad (**)$$

The right-hand side does not contain the required function. We put $y' = p$; then $y'' = p'$, and equation (**) becomes the first-order equation

$$p' = f(x, p).$$

This gives an expression for p in terms of x, and the solution is obtained by quadrature of the equation $y' = p$.

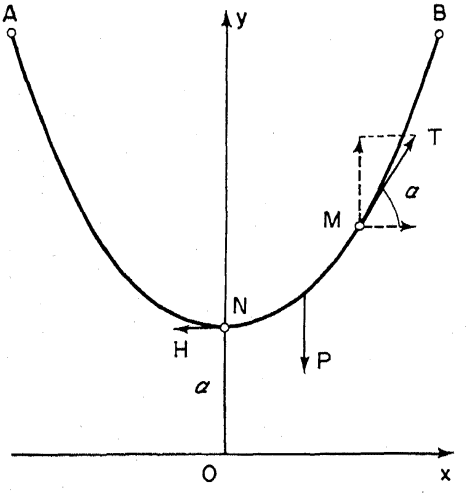

FIG. 73

A similar method is used for n-th order equations of the form $y^{(n)} = f(x, y^{(n-1)})$. On setting $y^{(n-1)} = p$, the problem reduces to integration of a first order equation and to subsequent integration of an equation of the form considered in I.

Example. Find the shape of a flexible, inextensible, homogeneous cord (chain) hanging from its two ends under the action of its own weight (Fig. 73).

We take as Oy a vertical straight line through the lowest point N of the curve; Ox is taken horizontally, at an as yet undetermined distance from point N.

Let M be an arbitrary point of the curve. In view of the equilibrium, the piece of cord NM can be regarded as a rigid body. It is subject to the action of three forces: the horizontal tension H, the

tension T acting at M and tangential to the curve at this point, and the weight P, equal to $s\delta$, where s is the length of cord NM and δ is the specific weight of the cord.

On resolving T into horizontal and vertical components and taking the equilibrium conditions into account, we clearly have

$$T\sin\alpha = s\delta, \quad T\cos\alpha = H.$$

We divide the first equation by the second:

$$\tan\alpha = \frac{\delta}{H}\, s.$$

Thus if $y = y(x)$ is the required equation of curve ANB, we have

$$y' = ks, \quad k = \frac{\delta}{H} = \text{const.}$$

We differentiate this equation with respect to x:

$$y'' = ks' = k\sqrt{1 + y'^2}.$$

We have arrived at an equation of form (**). Putting $y' = p$, we have $y'' = p'$, and

$$p' = k\sqrt{1 + p^2}, \quad \text{or} \quad \frac{dp}{\sqrt{1 + p^2}} = k\,dx,$$

whence

$$\ln\left(p + \sqrt{1 + p^2}\right) = kx + C_1.$$

At the point N, $x = 0$ and $p = y' = 0$ (since N is the lowest point of the cord). Thus $C_1 = 0$ and

$$p + \sqrt{1 + p^2} = e^{kx}, \quad \text{whence} \quad p = y' = \frac{1}{2}\left(e^{kx} - e^{-kx}\right).$$

Integration gives

$$y + C_2 = \frac{1}{2k}\left(e^{kx} + e^{-kx}\right).$$

We now choose distance $ON = 1/k$. Then $C_2 = 0$, and we get for the equation of the curve (*catenary*):

$$y = \frac{1}{2k}\left(e^{kx} + e^{-kx}\right).$$

This example enables us to see clearly the convenience of using the initial conditions to find in turn the values of the constants of

integration. If we first wrote down the general solution the working would be more cumbersome.

If we write a for $1/k$, the equation becomes

$$y = \frac{a}{2}\left(e^{\frac{x}{a}} + e^{-\frac{x}{a}}\right) = a \cosh \frac{x}{a}.$$

This is the familiar form of the catenary equation, which is so called because it gives the shape of a freely suspended heavy chain or cord.

If the conditions of the problem are changed and we seek the curve taken by a cord under the action of a horizontal homogeneous mass, the weight of the cord being negligible (*the problem of a suspension bridge*), the differential equation is considerably simplified. The result is a *parabola*. We suggest that the reader solve this problem for himself.

III. SECOND ORDER EQUATION OF THE FORM

$$y'' = f(y, y'). \qquad (***)$$

The right-hand side does not contain the independent variable. Again we put $y' = p$, but we now take p as a function of y. We get by differentiating this equation:

$$y'' = \frac{dp}{dx} = \frac{dp}{dy}\frac{dy}{dx} = \frac{dp}{dy}p.$$

Substitution in the original equation gives

$$\frac{dp}{dy}p = f(y, p),$$

i.e. a first order equation in p as a function of y. Having found p in terms of y, i.e. $p = \varphi(y)$, the required solution is got from the equation with separable variables:

$$\frac{dy}{dx} = p = \varphi(y), \quad \text{i.e.} \quad \frac{dy}{\varphi(y)} = dx.$$

The n-th order equation

$$y^{(n)} = f(y^{(n-2)}, y^{(n-1)})$$

is transformed by the substitution $y^{(n-2)} = z$ into a second order equation of form (***).

Various problems, including an important class of mechanical problems, lead to equations of form (***). Let the motion be due to the action of a force which can be expressed as a function depending explicitly only on the distance traversed and on the velocity (but not on time). The basic equation of mechanics now takes the form (***). Motions of this type include, in particular, vibrations due to the action of a force proportional to the distance, the resistive force being proportional to the velocity. Such vibrations are encountered extensively in engineering and physics.

Example. Let a body move along the Os axis under the action of a force with a fixed direction towards the origin (*central force*) and proportional to the displacement. The force defined by these conditions is an elementary example of a so-called elastic force. Since the force is equal to $\omega^2 s$, where ω^2 is a constant of proportionality, the equation of motion will be (neglecting the resistance)

$$s'' = -\omega^2 s.$$

We put $s' = p$; then $s'' = p\,dp/ds$. We have:

$$\frac{dp}{ds}\,p = -\omega^2 s,$$

or

$$p\,dp = -\omega^2 s\,ds.$$

Hence

$$p^2 = C - \omega^2 s^2.$$

Let $s = 0$ and $s' = p = A\omega$ at the initial instant $t = 0$. Now $C = A^2\omega^2$ and

$$p = \omega\sqrt{A^2 - s^2}.$$

Thus

$$\frac{ds}{dt} = \omega\sqrt{A^2 - s^2}.$$

Integration gives

$$s = A \sin \omega t.$$

We have obtained a *harmonic vibration* (cf. the example in I).

It is worth noting that all three types of second order equation: $y'' = f(x)$, $y'' = f(x, y')$, $y'' = f(y, y')$ are integrated with the aid of the same substitution $y' = p$.

201. Approximate Solutions. There are various methods for the approximate integration of second or higher order equations. The majority are of a specialized kind and relate to definite types of equation and definite problems. We shall only mention here a

simple graphical method for solving second order equations analogous to the graphical method for first order equations described in Sec. 196, and the widely used method of integration with the aid of series.

I. GRAPHICAL METHOD. Given the second order equation containing the second derivative explicitly:

$$y'' = f(x, y, y'), \qquad (*)$$

we wish to find its particular solution corresponding to the initial conditions $y|_{x=x_0} = y_0$, $y'|_{x=x_0} = y'_0$.

From the geometric point of view, the problem amounts to constructing the integral curve of the equation that passes through $M_0(x_0, y_0)$ and has a slope equal to y'_0 at this point.

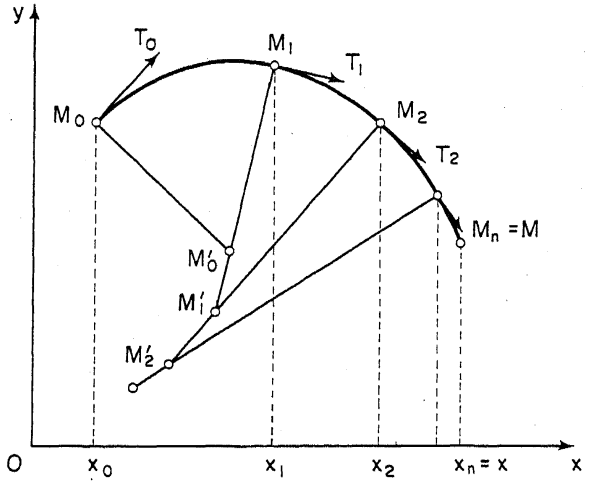

FIG. 74

We shall use a graphical method to find approximately the integral curve corresponding to the interval $[x_0, x]$.

We use the fact that the second order equation expresses a property of the integral curves relating to their slopes and curvatures. We divide $[x_0, x]$ into n sub-intervals by the points $x_0, x_1, x_2, \ldots, x_{n-1}, x_n = x$ (Fig. 74). We draw through the points of division straight lines parallel to Oy, then carry out the following sequence of operations.

We work out $f(x, y, y')$ at the point $M_0(x_0, y_0)$ in accordance with the initial conditions; in accordance with equation (*), $f(x_0, y_0, y'_0)$ gives the value of y'' at $x = x_0$, i.e. $y''(x_0) = y''_0$. Knowing this, we work out the radius of curvature R_0 of the integral curve at M_0:

$$R_0 = \left| \frac{(\sqrt{1 + y'^2_0})^3}{y''_0} \right|.$$

We draw a small straight arrow M_0T_0 from M_0 (Fig. 74) with a slope y'_0, then a segment $M_0M'_0$ perpendicular to it of length R_0 and in a sense cor-

responding to the sign of y_0''. Taking M_0' as centre, we describe a small arc of a circle of radius R_0, starting at the point M_0 and ending at the point M_1 of intersection of the circle with the straight line $x = x_1$. We next work out $f(x, y, y')$ at the point obtained $M_1(x_1, y_1)$ (the value $y' = y_1'$ is known at this point, since the tangent M_1T_1 is perpendicular to the "last" radius $M_0'M_1$); $f(x_1, y_1, y_1')$ gives the value of y'' at $x = x_1$, i.e. $y''(x_1) = y''$. Using the values obtained for y_1' and y_1'', we find the radius of curvature R_1 of the integral curve at point M_1:

$$R_1 = \left| \frac{(\sqrt{1 + y_1'^2})^3}{y_1''} \right|.$$

We next mark off along the radius M_1M_0' a segment M_1M_1', equal to the radius of curvature R_1. Taking the point M_1—the centre of curvature of the integral curve at M_1—as centre, we describe a small arc of a circle of radius R_1 as far as its intersection at M_2 with the next straight line $x = x_2$.

By proceeding in this way, we find in turn the points of the integral curve corresponding to the points of subdivision $x_3, x_4, \ldots$ of interval $[x_0, x]$ until we arrive at $M(x, y)$. The curve $M_0M_1M_2 \ldots M_n$ obtained *consists of arcs of circles and represents approximately the required integral curve.*

It may be said that the integral curve for a second order equation consists of arcs of circles in the same way as the integral curve for a first order equation is made up in Euler's method by pieces of tangents.

Obviously, the greater n and the smaller the maximum sub-interval $[x_i, x_{i+1}]$, the more accurate in general the result obtained by graphical integration of equation (*).

II. INTEGRATION WITH THE AID OF SERIES. The principal of applying power series to the integration of higher order equations is the same as in the case of first order equations. We suppose that the solution is expressible as a power series with undetermined coefficients which we proceed to find by using the equation. If the series obtained can be shown to be convergent it can be shown to give the required solution provided this is unique. There is usually no need for such proof.

The method of solution with the aid of power series is very widely used; it is in fact an extremely convenient and effective method for seeking approximate solutions of differential equations in a very simple form, viz. as polynomials in the independent variable. The method is often used for investigating linear equations (see § 4) of the second order with variable coefficients of the form

$$y'' + p_1(x)\, y' + p_2(x)\, y = 0$$

and —with a generalization of the power series —equations of the form

$$p_0(x)\, y'' + p_1(x)\, y' + p_2(x)\, y = 0,$$

which are of great importance in physics and engineering. In view of the specialized nature of all these methods we must leave them aside.

We shall just take one simple example for illustration:

$$y'' - xy = 0.$$

We shall find the solution of this equation with the initial conditions $y(0) = 0$, $y'(0) = 1$. Obviously, the solution must be sought as a series of the form

$$y = x + a_2x^2 + a_3x^3 + \cdots + a_nx^n + \cdots$$

Thus

$$y'' = 2a_2 + 3 \cdot 2a_3x + \cdots + n(n-1)\,a_nx^{n-2} + \cdots$$

i.e.

$$2a_2 + 3 \cdot 2a_3x + \cdots + n(n-1)\,a_nx^{n-2} + \cdots = x^2 + a_2x^3 + \cdots + a_{n-3}x^{n-2} + \cdots$$

We find on equating coefficients:

$$a_2 = 0, \quad a_3 = 0, \quad \cdots, \quad n(n-1)\,a_n = a_{n-3}.$$

Hence

$$a_4 = \frac{1}{3 \cdot 4}, \; a_5 = 0, \; a_6 = 0, \; a_7 = \frac{1}{3 \cdot 4 \cdot 6 \cdot 7}, \; a_8 = 0, \; a_9 = 0,$$

$$a_{10} = \frac{1}{3 \cdot 4 \cdot 6 \cdot 7 \cdot 9 \cdot 10},$$

and in general

$$a_{3m-1} = a_{3m} = 0, \quad a_{3m+1} = \frac{1}{3 \cdot 4 \cdot 6 \cdot 7 \ldots 3m(3m+1)}.$$

Thus

$$y = x + \frac{x^4}{3 \cdot 4} + \frac{x^7}{3 \cdot 4 \cdot 6 \cdot 7} + \cdots + \frac{x^{3m+1}}{3 \cdot 4 \cdot 6 \cdot 7 \ldots 3m(3m+1)} + \cdots$$

We can easily verify by using d'Alembert's test that this series is convergent throughout Ox and thus gives the required solution.

It may be mentioned in conclusion that the order of an equation has no effect on the method of solution with the aid of series.

4. Linear Equations

202. Homogeneous Equations. We turn to an important type of equation which is often encountered in all branches of applied mathematics, namely, the *linear equation*.

Definition. **A differential equation is said to be *linear* if it is of the first degree (linear) in the required function and its derivatives.**

A n-th order linear equation has the form

$$y^{(n)} + a_1 y^{(n-1)} + a_2 y^{(n-2)} + \cdots + a_{n-1} y' + a_n y = f, \qquad (*)$$

where coefficients $a_1, a_2, \ldots, a_{n-1}, a_n, f$ are functions of the independent variable x or constants (we assume that the coefficient of the highest order derivative is equal to 1*).

The function f is termed the *right-hand side* or the *free term*. If f is identically zero, equation (*) is said to be *linear without a right-hand side or free term*, or to be *homogeneous*. Otherwise (*) is called a *linear equation with right-hand side* (*or free term*), or is said to be *non-homogeneous*.

Continuity of the coefficients and free term of equation (*), which we shall always assume in future, guarantees that the conditions of the uniqueness and existence theorem are satisfied.

Linear equations of the first order ($n = 1$) were discussed in Sec. 194. As a rule, linear equations with $n > 1$ cannot be integrated with the aid of finite forms and quadratures. However, there is one class of equations (*) which is fairly wide from the point of view of applied mathematics for which complete integration is possible by solving algebraic equations and by quadratures. These are *linear equations with constant coefficients*. Before turning to these we shall mention some general theorems on linear equations required for the investigation of their solutions.

I. Structure of the general solution. We first require the concept of linear independence of two functions.

Definition. Two functions $y_1(x)$ and $y_2(x)$ are said to be linearly independent if their ratio is not constant, in other words, if constants k_1 and k_2 cannot be found such that the linear combination $k_1 y_1 + k_2 y_2$ is identically zero. (We assume here that at least one of the constants is non-zero, i.e. $k_1^2 + k_2^2 \neq 0$.)

* If the coefficient of the highest order derivative is not unity we can divide both sides of the equation by it for those values of x for which it differs from zero.

Thus if $k_1 y_1 + k_2 y_2 \neq 0$ no matter what constants k_1 and k_2 we choose (on condition that $k_1^2 + k_2^2 \neq 0$), functions y_1 and y_2 will be linearly independent. Whereas if

$$k_1 y_1 + k_2 y_2 = 0,$$

for certain constants k_1 and k_2 (let $k_2 \neq 0$), the ratio y_2/y_1 is constant ($= -k_1/k_2$), and one function is got by multiplying the other by a constant. Functions y_1 and y_2 are in this case *linearly dependent.*

We now prove a theorem which is fundamental to what follows.

OSTROGRADSKII'S THEOREM. **If y_1 and y_2 are two particular solutions of the second order linear homogeneous equation**

$$y'' + a_1 y' + a_2 y = 0 \tag{1}$$

we have

$$V(y_1, y_2) = y_1 y_2' - y_2 y_1' = V(y_{10}, y_{20})\, e^{-\int_{x_0}^{x} a_1 dx}, \tag{**}$$

where

$$y_{10} = y_1|_{x=x_0}, \quad y_{20} = y_2|_{x=x_0}.$$

Proof. Since y_1 and y_2 are solutions of equation (1), we have

$$y_1'' + a_1 y_1' + a_2 y_1 = 0,$$
$$y_2'' + a_1 y_2' + a_2 y_2 = 0.$$

On multiplying the first equation by y_2 and the second by y_1 and subtracting the first from the second, we get:

$$(y_1 y_2'' - y_2 y_1'') + a_1 (y_1 y_2' - y_2 y_1') = 0.$$

We observe that the expression in the second bracket is V and that in the first the derivative of V:

$$\frac{dV}{dx} = y_1' y_2' + y_1 y_2'' - y_2' y_1' - y_2 y_1'' = y_1 y_2'' - y_2 y_1''.$$

Thus

$$\frac{dV}{dx} + a_1 V = 0,$$

i.e.

$$\frac{dV}{V} = -a_1\, dx.$$

Integration from x_0 to x gives

$$\ln \frac{V}{V_0} = -\int_{x_0}^{x} a_1 dx,$$

where $V_0 = V|_{x=x_0} = V(y_{10}, y_{20})$. Hence

$$V = V_0 e^{-\int_{x_0}^{x} a_1 dx}.$$

This is what we had to prove.

Relationship (**) shows that V is identically zero if $V_0 = 0$, whilst if $V_0 \neq 0$, V vanishes nowhere, inasmuch as there is no x for which the second factor — the exponential function — vanishes.

The following is an important consequence of Ostrogradskii's theorem.

THEOREM. **If y_1 and y_2 are two linearly independent particular solutions of equation (1), we have a non-vanishing**

$$V(y_1, y_2) = y_1 y_2' - y_2 y_1'$$

whatever the value of x in the domain of continuity of the coefficients of the equation (this is called a *permissible* value).

Proof. Suppose, on the contrary, that $V|_{x=x_0} = V_0 = 0$, where x_0 is a permissible value of x; then by Ostrogradskii's theorem, $V = 0$ identically. But

$$\frac{d}{dx}\left(\frac{y_2}{y_1}\right) = \frac{y_1 y_2' - y_2 y_1'}{y_1^2} = \frac{V(y_1, y_2)}{y_1^2},$$

i.e. given our assumption, $d(y_2/y_1)/dx = 0$, so that y_2/y_1 is constant, which contradicts the assumed linear independence of y_1 and y_2.

The theorem is proved.

We can now state a fundamental proposition regarding the structure of the general solution of equation (1).

THEOREM. 1. **If y_1 and y_2 are linearly independent particular solutions of the equation**

$$y'' + a_1 y' + a_2 y = 0, \tag{1}$$

the general solution is equal to a linear combination of particular solutions y_1 and y_2 with arbitrary constants C_1 and C_2:

$$y = C_1 y_1 + C_2 y_2. \tag{2}$$

Proof. We show first of all that function (2) is a solution of equation (1) whatever the values of C_1 and C_2. We have

$$y' = C_1 y_1' + C_2 y_2', \quad y'' = C_1 y_1'' + C_2 y_2''. \tag{3}$$

We obtain on substituting expressions (2) and (3) in the left-hand side of equation (1):

$$C_1y_1'' + C_2y_2'' + a_1(C_1y_1' + C_2y_2') + a_2(C_1y_1 + C_2y_2)$$
$$= C_1(y_1'' + a_1y_1' + a_2y_1) + C_2(y_2'' + a_1y_2' + a_2y_2).$$

But the expressions in brackets are the result of substituting y_1 and y_2 respectively in the left-hand side of (1), and since they are solutions by hypothesis, these expressions must vanish identically, i.e. function (2) in fact satisfies equation (1).

We next verify that function (2) is in fact the general solution for arbitrary C_1 and C_2.

Given *any* initial conditions $y|_{x=x_0} = y_0$, $y'|_{x=x_0} = y_0'$, where $x = x_0$ is a permissible value of x, we show that, given our assumption regarding y_1 and y_2, C_1 and C_2 can be chosen such that function (2) satisfies these initial conditions. This will imply that (2) is in fact the general solution.

We must have:

$$y|_{x=x_0} = C_1y_{10} + C_2y_{20} = y_0,$$
$$y'|_{x=x_0} = C_1y_{10}' + C_2y_{20}' = y_0',$$

where

$$y_{10} = y_1(x_0), \qquad y_{10}' = y_1'(x_0), \qquad y_{20} = y_2(x_0), \qquad y_{20}' = y_2'(x_0).$$

The determinant of this system is V_0. Since $V_0 \neq 0$ by what has been proved, the system gives determinate finite values of C_1 and C_2:

$$C_1 = \frac{y_0y_{20}' - y_{20}y_0'}{y_{10}y_{20}' - y_{20}y_{10}'} = \frac{y_0y_{20}' - y_{20}y_0'}{V_0},$$

$$C_2 = \frac{y_{10}y_0' - y_0y_{10}'}{y_{10}y_{20}' - y_{20}y_{10}'} = \frac{y_{10}y_0' - y_0y_{10}'}{V_0}.$$

This is what we wanted to prove.

If particular solutions y_1 and y_2 are linearly dependent: $y_2/y_1 = k = \text{const}$, i.e. $y_2 = ky_1$, the function (2):

$$y = C_1y_1 + C_2y_2 = (C_1 + kC_2)y_1 = Cy_1$$

will in fact depend only on one arbitrary constant C (in view of the arbitrariness of C_1 and C_2 the constant $C_1 + kC_2$ can be regarded as a single arbitrary constant C). In this case function (2) does not yield the general solution.

Linear independent solutions are said to form a *fundamental system of solutions* of equation (1).

Thus, *to form the general solution of a second order linear homogeneous equation we need to know a fundamental system of solutions, i.e. any two linearly independent particular solutions.*

Example. Find the solution of

$$(x-1)y'' - xy' + y = 0$$

with the initial conditions $y|_{x=0} = 2$, $y'|_{x=0} = 1$.

It is not difficult simply to pick out two solutions of the equation. Functions $y = x$ and $y = e^x$ are soon seen to satisfy the equation. These particular solutions form a fundamental system, since e^x/x is not a constant; but neither x nor e^x satisfies the initial conditions.

We form the general solution:

$$y = C_1 x + C_2 e^x.$$

Hence

$$y' = C_1 + C_2 e^x.$$

On substituting the initial conditions in these equations we obtain a system of two equations for C_1 and C_2, viz.

$$2 = C_2,$$

$$1 = C_1 + C_2,$$

which give $C_2 = 2$, $C_1 = -1$. Thus the required solution is $y = -x + 2e^x$.

The above theory holds for n-th order linear homogeneous equations ($n > 2$).

Definition. We describe n functions $y_1, y_2, \ldots, y_n$ as linearly independent if it is impossible to choose constants $k_1, k_2, \ldots, k_n$ which are not all zero (i.e. $k_1^2 + k_2^2 + \cdots + k_n^2 \neq 0$) such that the linear combination

$$k_1 y_1 + k_2 y_2 + \cdots + k_n y_n$$

vanishes identically.

If a system of constants $k_1, k_2, \ldots, k_n$ exists for which

$$k_1 y_1 + k_2 y_2 + \cdots + k_n y_n = 0,$$

the functions $y_1, y_2, \ldots, y_n$ are *linearly dependent*, and any one of them (for which the coefficient k is non-zero) is given by a linear

combination of the remainder with constant coefficients. For instance, if $k_n \neq 0$, then

$$y_n = -\frac{k_1}{k_n} y_1 - \frac{k_2}{k_n} y_2 - \cdots - \frac{k_{n-1}}{k_n} y_{n-1}.$$

OSTROGRADSKII'S THEOREM. *If $y_1, y_2, \ldots, y_n$ are particular solutions of the equation*

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = 0, \tag{4}$$

we have

$$V = V_0 e^{-\int_{x_0}^{x} a_1 dx},$$

where V is the Wronskian of $y_1, y_2, \ldots, y_n$ and V_0 is its value at $x = x_0$.

The *Wronskian** of functions $y_1, y_2, \ldots, y_n$ is the determinant

$$V(y_1, y_2, \ldots, y_n) = \begin{vmatrix} y_1 & y_2 & \cdots & y_n \\ y_1' & y_2' & \cdots & y_n' \\ \cdots & \cdots & \cdots & \cdots \\ y_1^{(n-1)} & y_2^{(n-1)} & \cdots & y_n^{(n-1)} \end{vmatrix}.$$

The expression $V(y_1, y_2)$ used above for the case $n = 2$ is the second order Wronskian

$$V(y_1, y_2) = \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix}.$$

THEOREM. *If $y_1, y_2, \ldots, y_n$ are n linearly independent particular solutions of equation (4), the Wronskian $V(y_1, y_2, \ldots, y_n)$ does not vanish whatever the permissible value of x.*

We can use these theorems to extend theorem 1 to linear homogeneous equations of the n-th order.

THEOREM. *If $y_1, y_2, \ldots, y_n$ are n linearly independent particular solutions of equation (4), a linear combination of these solutions with arbitrary constant coefficients $C_1, C_2, \ldots, C_n$:*

$$y = C_1 y_1 + C_2 y_2 + \cdots + C_n y_n \tag{5}$$

is the general solution of (4).

If $y_1, y_2, \ldots, y_n$ are *linearly dependent*, at least one of the particular solutions is expressible in terms of the remaining $n - 1$, and function (5) will in fact depend on less than n arbitrary constants. It will not provide the general solution.

* I. G. WRONSKII (1778–1853), a celebrated Polish mathematician.

Linearly independent solutions of a linear n-th order equation are also said to form a *fundamental system of solutions.*

We shall not give the proof of the last three theorems in the general case*.

II. LOWERING THE ORDER. The following theorem often helps in finding the general solution.

THEOREM 2. **If one particular solution y_1 is known of the linear homogeneous second order equation**

$$y'' + a_1 y' + a_2 y = 0, \tag{1}$$

a particular solution y_2, linearly independent of y_1, may be found by quadratures of linear first order equations.

Proof. We use the substitution $y = y_1 u$, where u is an unknown function which must be chosen so that $y_1 u$ is a solution of (1). We have:

$$y' = y_1' u + y_1 u',$$
$$y'' = y_1'' u + 2y_1' u' + y_1 u''.$$

Substitution in the equation gives

$$y_1'' u + 2y_1' u' + y_1 u'' + a_1(y_1' u + y_1 u') + a_2 y_1 u = 0$$

or

$$(y_1'' + a_1 y_1' + a_2 y_1)u + [y_1 u'' + (2y_1' + a_1 y_1)u'] = 0.$$

Since y_1 is a solution of (1) by hypothesis, the expression in the first bracket is zero. We obtain the equation for u:

$$y_1 u'' + (2y_1' + a_1 y_1)u' = 0.$$

We put $u' = z$ and arrive at the first order equation with separable variables:

$$y_1 z' + (2y_1' + a_1 y_1)z = 0.$$

Knowing $z(\not\equiv 0)$, we find u by a single quadrature from the equation $u' = z$, then multiply to obtain $y_2 = y_1 u$. This solution is linearly independent of y_1, since $y_2/y_1 = u$ is a function which is not identically constant.

Example 1. We take the equation

$$(1 + 2x - x^2)y'' + (-3 + x^2)y' + (2 - 2x)y = 0.$$

Inspection shows that one solution is e^x. We put $y = e^x u$. We now get the equation for u:

$$e^x u'' + \left(2e^x + \frac{-3 + x^2}{1 + 2x - x^2} e^x\right) u' = 0$$

* See V.V. STEPANOV, *Course of Differential Equations* (*Kurs differentsial'nykh uravnenii*), 6th ed., Gost., 1953.

or, on cancelling e^x and writing z for u':

$$z' + \left(2 + \frac{-3 + x^2}{1 + 2x - x^2}\right) z = 0 .$$

Hence

$$\frac{dz}{z} = \frac{x^2 - 4x + 1}{-x^2 + 2x + 1} dx .$$

Integration gives

$$\ln z = -x + \ln(-x^2 + 2x + 1)$$

(we need any solution, so that the arbitrary constant can be given any desired value).

Further,

$$z = u' = e^{-x}(-x^2 + 2x + 1)$$

and further integration gives

$$u = \int e^{-x}(-x^2 + 2x + 1)dx = e^{-x}(x^2 - 1)$$

(we take zero for the arbitrary constant).

Thus

$$y_2 = y_1 u = x^2 - 1,$$

and the general solution of the given equation can be written as

$$y = C_1 e^x + C_2(x^2 - 1).$$

Example 2. The equation

$$x^2 y'' + xy' + (x^2 - n^2)y = 0 \quad (n = \text{const}),$$

known as *Bessel's equation*, is of great importance in several branches of physics. With $n = \frac{1}{2}$, this equation is satisfied by the function $y = \sin x/\sqrt{x}$. Knowing this, the reader will easily find the general solution of Bessel's equation with $n = \frac{1}{2}$.

Thus, *the problem of integrating a linear homogeneous second order equation reduces to finding any one particular solution.*

It can be shown as in the case of second order equations that a knowledge of one particular solution y_1 of a linear homogeneous n-th order equation enables us to reduce the problem of integrating the equation to the integration of a linear homogeneous equation of order $n - 1$ and to a subsequent quadrature. This is done in the same way, by replacing the required function y by a function u in accordance with the formula $y = y_1 u$.

In general, if $k (k < n)$ linearly independent solutions $y_1, y_2, \ldots, y_k$ of a linear homogeneous n-th order equation are known, integration of it reduces to integration of a linear homogeneous equation of order $n - k$ and to k quadratures.

For suppose we replace y by $y_1 u$, and u' by z. It is easily seen that $(y_2/y_1)'$ is a particular solution of the $(n-1)$-th order equation obtained for z. On replacing z in this equation by $(y_2/y_1)' v$, and v' by t, we arrive at an equation of order $n - 2$, and so on.

203. Non-homogeneous equations. We turn to second order linear *non-homogeneous equations*:

$$y'' + a_1 y' + a_2 y = f. \tag{1}$$

If we take 0 instead of the free term f, we get the *homogeneous* equation

$$y'' + a_1 y' + a_2 y = 0, \tag{2}$$

which is said to correspond to the given non-homogeneous equation.

I. We prove the following fundamental theorem on non-homogeneous equations.

THEOREM 1. **The general solution of a non-homogeneous equation can be written as the sum of a particular solution of this equation and the general solution of the corresponding homogeneous equation.**

Proof. Let $\bar{y}$ denote a particular solution of equation (1), and Y the general solution of equation (2); we put

$$y = \bar{y} + Y. \tag{3}$$

We substitute function (3) in the left-hand side of equation (1). Since

$$y' = \bar{y}' + Y', \quad y'' = \bar{y}'' + Y'',$$

we get

$$\bar{y}'' + Y'' + a_1(\bar{y}' + Y') + a_2(\bar{y} + Y)$$
$$= (\bar{y}'' + a_1 \bar{y}' + a_2 \bar{y}) + (Y'' + a_1 Y' + a_2 Y).$$

In view of the fact that $\bar{y}$ is a solution of equation (1), $\bar{y}'' + a_1 \bar{y}' + a_2 \bar{y}$ is identically equal to function f; the expression $Y'' + a_1 Y' + a_2 Y$ is identically zero, since Y is a solution of equation (2). Thus function (3) turns equation (1) into an identity, in other words, it is a solution. But it depends on two arbitrary constants C_1, C_2 (the second term Y depends on them), which can always be chosen so as to satisfy any initial conditions; the proof is as

for homogeneous equations. Function (3) is therefore the general solution of equation (1).

This is what we wished to prove.

Thus to find the general solution of a second order non-homogeneous linear equation we only need to know any one particular solution and the general solution of the corresponding homogeneous equation, i.e. in the last analysis, *one particular solution of the non-homogeneous and one particular solution of the homogeneous equation.*

Example. Given

$$(1 + 2x - x^2)y'' + (-3 + x^2)y' + (2 - 2x)y = -x^2 + 2x - 3$$

we want to find the solution with initial conditions

$$y|_{x=0} = y'|_{x=0} = 0.$$

We pick out the particular solution $\bar{y} = x$, though it does not satisfy the initial conditions. The general solution of the corresponding homogeneous equation is known (see Sec. 202); it enables the general solution of the non-homogeneous equation to be written:

$$y = x + C_1 e^x + C_2(x^2 - 1).$$

We use the initial conditions to find the values of the arbitrary constants: $C_1 = C_2 = -1$. The required solution is

$$y = 1 + x - x^2 - e^x.$$

A similar theorem may be proved for n-th order non-homogeneous equations.

THEOREM. *If we know a particular solution y of the non-homogeneous equation*

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1}y' + a_n y = f$$

and the general solution Y of the corresponding homogeneous equation, the general solution y of the non-homogeneous equation is equal to the sum of $\bar{y}$ and Y:

$$y = \bar{y} + Y.$$

II. If we know the general solution Y of a homogeneous equation, a particular solution of any corresponding non-homogeneous equation can in fact always be found with the aid of quadratures. There are various ways of doing this. We shall give the most extensively used method — that of variation of the arbitrary constants, due to Lagrange.

THEOREM 2. **A particular solution of non-homogeneous linear equation (1) can be got simply by replacing arbitrary constants C_1 and C_2 in the expression for the general solution of homogeneous equation (2):**

$$C_1 y_1 + C_2 y_2$$

by functions of the independent variable whose derivatives C_1' and C_2' satisfy the following system of linear algebraic equations:

$$C_1' y_1 + C_2' y_2 = 0,$$
$$C_1' y_1' + C_2' y_2' = f.$$

Proof. We try to choose as C_1 and C_2 functions of the independent variable x such that their linear combination with the particular solutions y_1 and y_2 of homogeneous equation (2) satisfies non-homogeneous equation (1).

Differentiation of

$$y = C_1 y_1 + C_2 y_2$$

gives us

$$y' = C_1 y_1' + C_2 y_2' + (C_1' y_1 + C_2' y_2).$$

Since functions C_1 and C_2 have to be chosen, we can arrange one relationship between them as desired. We put

$$C_1' y_1 + C_2' y_2 = 0.$$

Now:

$$y' = C_1 y_1' + C_2 y_2',$$

whence we find by means of further differentiation:

$$y'' = C_1 y_1'' + C_2 y_2'' + (C_1' y_1' + C_2' y_2').$$

Substitution of y, y', y'' in the left-hand side of (1) gives

$$(C_1 y_1'' + C_2 y_2'') + (C_1' y_1' + C_2' y_2') + a_1 (C_1 y_1' + C_2 y_2') +$$
$$+ a_2 (C_1 y_1 + C_2 y_2) = C_1 (y_1'' + a_1 y_1' + a_2 y_1) +$$
$$+ C_2 (y_2'' + a_1 y_2' + a_2 y_2) + (C_1' y_1' + C_2' y_2').$$

The expressions in the first and second brackets on the right-hand side vanish identically, since y_1 and y_2 are particular solutions of the homogeneous equation. Thus the necessary and sufficient condition for the function $C_1 y_1 + C_2 y_2$ satisfying the condition $C_1' y_1 + C_2' y_2 = 0$ to be a solution of (1) is that also

$$C_1' y_1' + C_2' y_2' = f.$$

We have thus arrived at the two equations

$$\left.\begin{aligned} C_1'y_1 + C_2'y_2 &= 0, \\ C_1'y_1' + C_2'y_2' &= f, \end{aligned}\right\} \tag{4}$$

from which C_1' and C_2' can be found uniquely, inasmuch as $V(y_1, y_2) = y_1y_2' - y_2y_1' \neq 0$; then C_1 and C_2 can be found by quadratures. If the arbitrary constants are brought in when integrating C_1' and C_2', we at once obtain the general solution of the non-homogeneous equation.

Example. We solve the non-homogeneous equation

$$x^2y'' - 2xy' + 2y = 2x^3.$$

We take the corresponding homogeneous equation

$$x^2y'' - 2xy' + 2y = 0.$$

Obviously one solution is $y_1 = x$. We find a second particular solution by the familiar method: $y_2 = x^2$.

The general solution of the homogeneous equation is

$$Y = C_1x + C_2x^2.$$

We use the method of variation of the arbitrary constants to find a particular solution of the non-homogeneous equation. We take C_1 and C_2 as functions of x such that $C_1x + C_2x^2$ satisfies the given equation; we now get two linear algebraic equations for C_1' and C_2':

$$C_1'x + C_2'x^2 = 0,$$

$$C_1' + 2C_2'x = 2x$$

(we take $2x$, and not $2x^3$ on the right-hand side of the second equation because equations (4) were deduced on the assumption that the coefficient of y'' is unity).

We find that

$$C_1' = -2x, \quad C_2' = 2.$$

Hence

$$C_1 = -x^2, \quad C_2 = 2x.$$

The function $y = -x^2 \cdot x + 2x \cdot x^2 = x^3$ must therefore be a solution of the non-homogeneous equation. This result may be verified by direct substitution. The general solution of the given non-homogeneous equation takes the form

$$y = x^3 + C_1x + C_2x^2.$$

If we take $C_1 = -x^2 + D_1$ and $C_2 = 2x + D_2$, where D_1 and D_2 are arbitrary constants, we get the general solution directly:

$$y = (-x^2 + D_1)x + (2x + D_2)x^2 = x^3 + D_1 x + D_2 x^2.$$

Thus *to integrate a non-homogeneous linear equation of the second order we only need to find one particular solution of the corresponding homogeneous equation.*

Given a knowledge of this, a second particular solution is found, linearly independent of the first (Sec. 202, THEOREM 2), and the general solution formed (Sec. 202, THEOREM 1); a particular solution of the non-homogeneous equation is then found from this by the method of variation of the arbitrary constants; and finally, the general solution is obtained by adding this particular solution to the general solution of the corresponding homogeneous equation (the last two steps can be replaced by one).

Variation of the arbitrary constants can also be used in the case of a linear equation of any order n. The following theorem is obtained precisely as above.

THEOREM. *The function*

$$C_1 y_1 + C_2 y_2 + \cdots + C_n y_n,$$

where $y_1, y_2, \ldots, y_n$ *is a fundamental system of solutions of the equation*

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_n y = 0,$$

is a solution of the non-homogeneous equation

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_n y = f,$$

if $C_1, C_2, \ldots, C_n$ *are functions of the independent variable whose derivatives* $C_1', C_2', \ldots, C_n'$ *satisfy the following system on n linear algebraic equations:*

$$\begin{aligned} C_1' y_1 + C_2' y_2 + \cdots + C_n' y_n &= 0, \\ C_1' y_1' + C_2' y_2' + \cdots + C_n' y_n' &= 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ C_1' y_1^{(n-1)} + C_2' y_2^{(n-1)} + \cdots + C_n' y_n^{(n-1)} &= f. \end{aligned}$$

5. Linear Equations with Constant Coefficients

204. Homogeneous Equations. We now show that the general solution of a homogeneous linear equation with *constant* coefficients may be found in the finite form *without the aid of quadratures.*

Let us take the linear homogeneous second order equation

$$y'' + a_1 y' + a_2 y = 0, \tag{1}$$

where a_1 and a_2 are constants.

Our problem consists in picking out at least one particular solution of this equation. We try to satisfy it with a function of the form $y = e^{rx}$ ($r = \text{const}$).

We have:

$$y' = re^{rx}, \quad y'' = r^2 e^{rx}.$$

We must therefore have identically

$$e^{rx}(r^2 + a_1 r + a_2) = 0$$

or

$$r^2 + a_1 r + a_2 = 0. \tag{2}$$

It is clear from this that e^{rx} in fact satisfies the differential equation if r satisfies the algebraic quadratic equation (2).

We call (2) the characteristic equation.

This is formed from the given differential equation (1) by replacing each derivative of the required function ($y = y^0, y', y''$) by r to a power equal to the order of the derivative ($1 = r^0, r, r^2$).

The search for particular solutions of linear homogeneous second order equations with constant coefficients thus reduces to the algebraic operation of solving quadratic equations.

Three possible cases must be distinguished as regards the roots r_1 and r_2 of the characteristic equation (we shall always assume that coefficients a_1 and a_2 are real numbers): (1) r_1 and r_2 are real and distinct, $r_1 \neq r_2$; (2) r_1 and r_2 are complex conjugates $r_1 = \alpha + \beta i$, $r_2 = \alpha - \beta i$; (3) r_1 and r_2 are real and equal, $r_1 = r_2$ (r_1 is the square root of equation (2)).

We shall consider each of these cases separately.

I. THE ROOTS OF THE CHARACTERISTIC EQUATION ARE REAL AND DISTINCT. In this case both roots (r_1 and r_2) can be taken as the index r of e^{rx}, and we obtain two solutions of equation (1) immediately: $e^{r_1 x}$ and $e^{r_2 x}$. Since the ratio $e^{r_1 x}/e^{r_2 x} = e^{(r_1 - r_2)x}$ is not constant by virtue of the inequality $r_1 \neq r_2$, these solutions are linearly independent.

The general solution is given in this case by

$$y = C_1 e^{r_1 x} + C_2 e^{r_2 x}.$$

Example.

$$y'' - y' - 2y = 0.$$

We form the characteristic equation:

$$r^2 - r - 2 = 0.$$

Its roots are

$$r_1 = 2, \quad r_2 = -1.$$

The general solution is therefore

$$y = C_1 e^{2x} + C_2 e^{-x}.$$

II. THE ROOTS OF THE CHARACTERISTIC EQUATION ARE COMPLEX CONJUGATES. This is the same as the previous case from the formal point of view. The equation is satisfied by $e^{(\alpha+\beta i)x}$ and by $e^{(\alpha-\beta i)x}$, these being linearly independent solutions since their ratio $e^{2\beta i x}$ is not a constant. Thus the general solution could be written as

$$y = C_1 e^{(\alpha+\beta i)x} + C_2 e^{(\alpha-\beta i)x}.$$

However, when an equation is specified in the domain of real numbers, we usually try to keep its solution in this domain and not go over to complex numbers.

We show how this is achieved in the present case. The following must first be pointed out.

If a given function $y = y_1 + iy_2$, where y_1 and y_2 are real functions of the independent variable x, satisfy equation (1) *(with real, but not necessarily constant coefficients), each separate function must be a solution of* (1).

We have, in fact:

$$(y_1'' + iy_2'') + a_1(y_1' + iy_2') + a_2(y_1 + iy_2) = 0,$$

or

$$(y_1'' + a_1 y_1' + a_2 y_1) + i(y_2'' + a_1 y_2' + a_2 y_2) = 0;$$

but a complex expression can only vanish identically when both its real and imaginary parts each vanish identically; hence

$$y_1'' + a_1 y_1' + a_2 y_1 = 0$$

and

$$y_2'' + a_1 y_2' + a_2 y_2 = 0.$$

This means that y_1 and y_2 satisfy equation (1). Now let the characteristic equation have complex roots

$$r_1 = \alpha + \beta i, \quad r_2 = \alpha - \beta i.$$

In this case

$$e^{(\alpha+\beta i)x} = e^{\alpha x} e^{\beta i x}$$

is a solution of equation (1). We have by Euler's formula (Sec. 135):

$$e^{\alpha x} e^{\beta i x} = e^{\alpha x}(\cos\beta x + i \sin\beta x) = e^{\alpha x}\cos\beta x + i e^{\alpha x}\sin\beta x,$$

i.e. by what has just been proved, both the functions $e^{\alpha x}\cos\beta x$ and $e^{\alpha x}\sin\beta x$ are solutions of the equation. Their linear independence is obvious. We can write the general solution:

$$y = C_1 e^{\alpha x}\cos\beta x + C_2 e^{\alpha x}\sin\beta x = e^{\alpha x}(C_1\cos\beta x + C_2\sin\beta x).$$

If we neglect the sign, the second root $r_2 = \alpha - \beta i$ gives the same particular solutions:

$$e^{\alpha x}\cos\beta x \quad \text{and} \quad -e^{\alpha x}\sin\beta x.$$

The general solution formed from these solutions will have the same form.

Thus the general solution of equation (1) can be written in the present case as

$$y = e^{\alpha x}(C_1\cos\beta x + C_2\sin\beta x),$$

where α is the real and β the imaginary part of the complex root of characteristic equation (2).

Example 1. $y'' - 4y' + 13y = 0$. We write down the characteristic equation

$$r^2 - 4r + 13 = 0.$$

Its roots are

$$r_1 = 2 + 3i, \quad r_2 = 2 - 3i.$$

The general solution is therefore

$$y = e^{2x}(C_1\cos 3x + C_2\sin 3x).$$

Example 2. $y'' + \omega^2 y = 0$ (ω = const).

We have already encountered this equation in Sec. 200, III, where it was integrated by a special method. We now find the general solution by the present method.

The characteristic equation

$$r^2 + \omega^2 = 0$$

gives

$$r_1 = \omega i, \quad r_2 = -\omega i.$$

Therefore $$y = C_1 \cos \omega x + C_2 \sin \omega x.$$

If we take the initial conditions of the example of Sec. 200:

$$y|_{x=0} = 0, \quad y'|_{x=0} = A\omega,$$

we have

$$C_1 = 0, \quad C_2 = A,$$

and we arrive at the result obtained previously:

$$y = A \sin \omega x.$$

III. THE ROOTS OF THE CHARACTERISTIC EQUATION ARE REAL AND COINCIDENT. When $r_1 = r_2$ we can only obtain one particular solution of equation (1) directly, viz $y_1 = e^{r_1 x}$. We employ the general rule of Sec. 202 (THEOREM 2) to find a second solution, linearly independent of y_1. We put $y = y_1 u$ and arrive at the equation

$$(2y_1' + a_1 y_1)u' + y_1 u'' = 0.$$

But the expression in brackets vanishes. For,

$$2y_1' + a_1 y_1 = 2r_1 e^{r_1 x} + a_1 e^{r_1 x} = e^{r_1 x}(2r_1 + a_1),$$

whilst by the property of quadratic equations the sum $2r_1$ of the roots is equal to the coefficient of the first power of r taken with the opposite sign, i.e. is equal to $-a_1$, so that $2r_1 + a_1 = 0$.

We thus have

$$y_1 u'' = 0, \quad \text{i.e.} \quad u'' = 0.$$

Any function can be taken that satisfies this equation; we choose the simplest: $u = x$. Hence $y_2 = xe^{r_1 x}$ is a second particular solution of equation (1).

Thus the general solution of equation (1) is given in the present case by

$$y = C_1 e^{r_1 x} + C_2 x e^{r_1 x} = e^{r_1 x}(C_1 + C_2 x).$$

Example.

$$16y'' - 24y' + 9y = 0.$$

The characteristic equation

$$16r^2 - 24r + 9 = 0$$

has a double root

$$r_1 = r_2 = \frac{3}{4},$$

so that the general solution is

$$y = e^{\frac{3}{4}x}(C_1 + C_2 x).$$

All in all, integration of a second order linear homogeneous equation with constant coefficients is carried out without any quadratures and only involves the solution of algebraic quadratic equations.

A similar result holds for linear homogeneous equations with constant coefficients of orders greater than two:

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = 0, \tag{3}$$

where $a_1, a_2, \ldots, a_n$ are constants.

We omit the proofs, which will involve the reader in no particular difficulty if he wishes to deduce them for himself.

As in the case of second order equations, the function e^{rx} is a particular solution of equation (3) if r is a root of the characteristic equation

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0.$$

Each root of the characteristic equation yields one particular solution of equation (3), although the cases of complex roots and multiple roots require special consideration if we are to find n linearly independent (and at the same time real) solutions of (3) (a fundamental system). We shall state the final results (using capital letters C and D to denote arbitrary constants):

(1) For every simple root r of the characteristic equation there is a corresponding term of the form Ce^{rx} in the general solution.

(2) For every pair of simple complex conjugate roots $r_1 = \alpha + \beta i$, $r_2 = \alpha - \beta i$ there is a corresponding term in the general solution:

$$e^{\alpha x}(C_1 \cos \beta x + C_2 \sin \beta x).$$

(3) For every k-ple root there is a corresponding term in the general solution:

$$e^{rx}(C_1 + C_2 x + \cdots + C_k x^{k-1}).$$

(4) For every pair of t-ple complex conjugate roots $r_1 = \alpha + \beta i$, $r_2 = \alpha - \beta i$ there is a corresponding term in the general solution:

$$e^{\alpha x}[(C_1 + C_2 x + \cdots + C_t x^{t-1}) \cos \beta x + (D_1 + D_2 x + \cdots + D_t x^{t-1}) \sin \beta x].$$

The total number of arbitrary constants is equal to the order of the equation, and the sum of the above terms yields the general solution.

Thus, *the solution of linear homogeneous equations with constant coefficients amounts simply to solving algebraic equations.*

Example.

$$y^{(V)} + y^{(IV)} + 2y''' + 2y'' + y' + y = 0.$$

The characteristic equation

$$r^5 + r^4 + 2r^3 + 2r^2 + r + 1 = 0$$

is easily seen to have the root $r = -1$; after division by $r + 1$ the equation becomes

$$r^4 + 2r^2 + 1 = 0,$$

i.e.

$$(r^2 + 1)^2 = 0.$$

We thus have:

$$r_1 = -1, \quad r_2 = r_3 = i, \quad r_4 = r_5 = -i.$$

The general solution of the differential equation is therefore

$$y = C_1 e^{-x} + (C_2 + C_3 x) \cos x + (C_4 + C_5 x) \sin x.$$

205. Non-homogeneous Equations. Suppose now that we are given the second order non-homogeneous linear equation with constant coefficients

$$y'' + a_1 y' + a_2 y = f(x), \tag{1}$$

where a_1 and a_2 are constants, and $f(x)$ is a given function of the independent variable x. We already know the principle underlying the method of solution: having found (without the aid of quadratures) the general solution of the corresponding homogeneous equation, we use the method of variation of the arbitrary constants (by means of quadratures) to find a particular solution of equation (1); we then form its general solution (see Sec. 203) and finally obtain the particular solution needed.

This method sometimes becomes rather unwieldy in practice. Moreover, when the free term $f(x)$ takes certain special forms it is possible to point to very simple ways of finding particular solutions of equation (1) *without the aid of quadratures.*

We shall deal with finding particular solutions of (1) in the cases when (1) $f(x) = P(x)\, e^{\alpha x}$, where $P(x)$ is a polynomial and $\alpha = \text{const}$

and in particular, zero; (2) $f(x) = P(x)\, e^{\alpha x} \cos \beta x$ (or $f(x) = P(x)\, e^{\alpha x} \sin \beta x$), where $P(x)$ is a polynomial, and α and β are constants.

These are the cases of most interest for applications.

We first mention a proposition of great practical importance.

Let the free term of equation (1) *be equal to the sum of two functions:* $f(x) = f_1(x) + f_2(x)$, *whilst* y_1 *and* y_2 *are solutions of the equation with the same left-hand side as* (1) *but with right-hand sides equal to* $f_1(x)$ *and* $f_2(x)$ *respectively; then* $y_1 + y_2$ *is a solution of equation* (1).

For,

$$\begin{aligned}(y_1'' + y_2'') + a_1(y_1' + y_2') + a_2(y_1 + y_2) &\\ = (y_1'' + a_1 y_1' + a_2 y_1) + (y_2'' + a_1 y_2' + a_2 y_2) &\\ = f_1(x) + f_2(x) = f(x).&\end{aligned}$$

This proposition holds for linear equations of any order with any coefficients.

I. $f(x) = P(x)\, e^{\alpha x}$.

We show that, if α is not a root of the characteristic equation, equation (1) is satisfied by a function of the same form as $f(x)$, namely $y = Q(x)\, e^{\alpha x}$, where $Q(x)$ is a polynomial of the same degree as $P(x)$ (the degree will be denoted by m).

We take a polynomial $Q(x)$ with as yet undetermined coefficients (there will be $m + 1$ of them) and replace y by the function $Q(x)\, e^{\alpha x}$ on the left-hand side of (1); we get

$$\begin{aligned}Q''(x)\, e^{\alpha x} + 2\alpha Q'(x)\, e^{\alpha x} + \alpha^2 Q(x)\, e^{\alpha x} + &\\ + a_1[Q'(x)\, e^{\alpha x} + \alpha Q(x)\, e^{\alpha x}] + a_2 Q(x)\, e^{\alpha x} &\\ = e^{\alpha x}[Q''(x) + (2\alpha + a_1) Q'(x) + (\alpha^2 + a_1\alpha + a_2) Q(x)] = P(x)\, e^{\alpha x},&\end{aligned}$$

$$\text{i.e. } Q''(x) + (2\alpha + a_1) Q'(x) + (\alpha^2 + a_1\alpha + a_2) Q(x) = P(x). \quad (2)$$

It follows from this that function $Q(x)\, e^{\alpha x}$ will be a solution of equation (1) if the coefficients of polynomial $Q(x)$ can be chosen so that equation (2) is satisfied identically. But since the expression on the left-hand side is a polynomial of degree m (because $\alpha^2 + a_1\alpha + a_2 \neq 0$), this can in fact be accomplished by equating coefficients of like powers of x on the two sides of equation (2); we get a system of $m + 1$ equations in the $m + 1$ unknown coefficients of $Q(x)$, whence the coefficients can be found.

If α is a root of the characteristic equation, the left-hand side of equation (2) contains a polynomial of degree $m - 1$ instead of

m (if α is a simple root), or of degree $m-2$ (if α is a double root, in which case $2\alpha + a_1 = 0$ as well as $\alpha^2 + a_1\alpha + a_2 = 0$).

Obviously, to satisfy the equation by a function of the form $Q(x)\,e^{\alpha x}$ in this case, we only need to increase the degree of $Q(x)$ (by one or two) without, however, increasing the number of its coefficients. This can be achieved by multiplying the m-th degree polynomial by x or by x^2, depending on whether α is a simple or double root of the characteristic equation.

Thus:

Equation (1) with the right-hand side $P(x)e^{\alpha x}$ has the particular solution

$$y = x^k Q(x) e^{\alpha x},$$

where $Q(x)$ is a polynomial of the same degree as $P(x)$ and k is the multiplicity of α as a root of the characteristic equation (if α is not a root, we take k equal to zero).

This assertion is applicable without modification to linear equations of any order with constant coefficients.

Knowing in advance the form of the solution, we then find the polynomial $Q(x)$ by the method of undetermined coefficients.

Example. $y'' - 2y' + y = 1 + x + 2(3x^2 - 2)\,e^x$.

We split the free term into two and consider the two equations

$$y'' - 2y' + y = 1 + x,$$

$$y'' - 2y' + y = 2(3x^2 - 2)\,e^x.$$

In the first equation $m = 1$ and $\alpha = 0$; and $\alpha = 0$ is not a root of the characteristic equation $r^2 - 2r + 1 = 0$. We put $y = ax + b$ and substitute in the equation; we get

$$-2a + ax + b = 1 + x,$$

whence $a = 1$, $b = 3$. Thus $y = x + 3$ is a particular solution of the first equation.

In the second equation $m = 2$ and $\alpha = 1$, this latter being a double root of the characteristic equation. We seek the solution in the form

$$y = x^2(ax^2 + bx + c)\,e^x.$$

We have:

$$y' = [ax^4 + (4a + b)x^3 + (3b + c)x^2 + 2cx]e^x,$$

$$y'' = [ax^4 + (8a + b)x^3 + (12a + 6b + c)x^2 + (6b + 4c)x + 2c]e^x.$$

Substitution in the equation leads to

$$12ax^2 + 6bx + 2c = 2(3x^2 - 2),$$

from which we find by equating coefficients: $a = \frac{1}{2}$, $b = 0$, $c = -2$. Thus $y = x^2(\frac{1}{2}x^2 - 2)\,e^x$ is a particular solution of the second equation.

Our particular solution of the given equation is therefore

$$y = x + 3 + x^2\left(\frac{1}{2}x^2 - 2\right)e^x.$$

The general solution is obviously

$$y = x + 3 + x^2\left(\frac{1}{2}x^2 - 2\right)e^x + (C_1 + C_2x)e^x.$$

II. $f(x) = P(x)\,e^{\alpha x}\cos\beta x$ (or $f(x) = P(x)\,e^{\alpha x}\sin\beta x$).

This case is the same as the previous one if we introduce complex numbers and pass from the trigonometric to exponential functions. We transform the free term in accordance with Euler's formula (Sec. 135):

$$P(x)e^{\alpha x}\cos\beta x = \frac{1}{2}P(x)e^{\alpha x}(e^{\beta xi} + e^{-\beta xi})$$

$$= \frac{1}{2}P(x)e^{(\alpha+\beta i)x} + \frac{1}{2}P(x)e^{(\alpha-\beta i)x}.$$

As in case I, we can seek the solution as

$$Q_1(x)e^{(\alpha+\beta i)x} + Q_2(x)e^{(\alpha-\beta i)x},$$

where $Q_1(x)$ and $Q_2(x)$ are polynomials of the same degree as $P(x)$, it being necessary to add the factor x if $\alpha \pm \beta i$ are roots of the characteristic equation. On returning to trigonometric functions it can be shown (though we shall not dwell on this) that equation (1) admits of a particular solution in a form which no longer contains complex numbers:

$$y = e^{\alpha x}[R_1(x)\cos\beta x + R_2(x)\sin\beta x],$$

where $R_1(x)$ and $R_2(x)$ are polynomials of the same degree as $P(x)$ with real coefficients.

The same is true with $f(x) = P(x)\,e^{\alpha x}\sin\beta x$.

It is clear that in the more general case when

$$f(x) = e^{\alpha x}[P_1(x)\cos\beta x + P_2(x)\sin\beta x],$$

the equation has a particular solution of the same form except that the degrees of polynomials $R_1(x)$ and $R_2(x)$ have to be taken equal to the higher of the degrees of polynomials $P_1(x)$ and $P_2(x)$.

Thus:

Equation (1) with right-hand side equal to $e^{\alpha x}[P_1(x)\cos\beta x + P_2(x)\sin\beta x]$ has a particular solution

$$y = x^k e^{\alpha x}[R_1(x)\cos\beta x + R_2(x)\sin\beta x]$$

where $R_1(x)$, $R_2(x)$ are polynomials of degree equal to the higher of the degrees of polynomials $P_1(x)$, $P_2(x)$, and k is the multiplicity with which $\alpha \pm \beta i$ appears as a root of the characteristic equation (if $\alpha \pm \beta i$ is not a root of the characteristic equation we take k equal to zero).

This statement applies to linear equations with constant coefficients of any order.

Knowing the form of the solution, we find the actual polynomials $R_1(x)$ and $R_2(x)$ by the method of undetermined coefficients.

Example. $y'' + y = 4x\sin x$.

Here $\alpha = 0$, $\beta = 1$ and $\pm i$ are roots of the characteristic equation $r^2 + 1 = 0$. We therefore seek the particular solution as

$$y = x[(ax + b)\cos x + (a_1 x + b_1)\sin x],$$

where a, b, a_1, b_1 are undetermined coefficients which have to be chosen so that this expression for y satisfies the equation.

We have

$$y'' = [-ax^2 + (4a_1 - b)x + (2a + 2b_1)]\cos x + [-a_1 x^2 - (4a + b_1)x + (2a_1 - 2b)]\sin x.$$

Substitution in the equation gives

$$[2a_1 x + (a + b_1)]\cos x + [-2ax + (a_1 - b)]\sin x = 2x\sin x.$$

This will be an identity only when

$$2a_1 = 0, \quad a + b_1 = 0, \quad -2a = 2, \quad a_1 - b = 0.$$

Hence

$$a = -1, \quad b = 0, \quad a_1 = 0, \quad b_1 = 1.$$

We thus get the particular solution

$$y = x(\sin x - x\cos x).$$

The general solution is given by

$$y = x(\sin x - x\cos x) + C_1\cos x + C_2\sin x$$
$$= (C_1 - x^2)\cos x + (C_2 + x)\sin x.$$

206. General Formula for the Solution of a Non-homogeneous Equation. If the free term has neither of forms I or II of Sec. 205, we must have recourse to the method of variation of the arbitrary constants or to some other general method for finding a particular solution of the *non-homogeneous* equation

$$y'' + a_1 y' + a_2 y = f(x). \tag{1}$$

We shall describe one such general method: it consists in using fairly simple working to find a particular solution of the non-homogeneous equation from a knowledge of a particular solution of the corresponding homogeneous equation.

We in fact find the particular solution $y(x)$ of equation (1) satisfying the initial conditions

$$y|_{x=0} = y'|_{x=0} = 0,$$

with the aid of the particular solution $\psi(x)$ of the corresponding homogeneous equation satisfying the conditions

$$\psi(0) = 0, \quad \psi'(0) = 1.$$

We take $x - t$ instead of x as the argument of function ψ and form the expression

$$I(x) = \int_0^x \psi(x - t)\, f(t)\, dt.$$

We show by direct substitution of $I(x)$ for y in equation (1) that $I(x)$ is in fact the required particular solution $y(x)$ of equation (1).

We have by the rule (Sec. 180) for differentiation of an integral with respect to a parameter (which is here the variable x):

$$I'(x) = \psi(x - t)\, f(t)|_{t=x} + \int_0^x \psi'(x - t)\, f(t)\, dt$$
$$= \psi(0)\, f(x) + \int_0^x \psi'(x - t)\, f(t)\, dt.$$

Since $\psi(0) = 0$ by hypothesis, we have

$$I'(x) = \int_0^x \psi'(x - t)\, f(t)\, dt.$$

Similarly,

$$I''(x) = \psi'(x-t)\,f(t)|_{t=x} + \int_0^x \psi''(x-t)\,f(t)\,dt$$

$$= \psi'(0)\,f(x) + \int_0^x \psi''(x-t)\,f(t)\,dt.$$

Since $\psi'(0) = 1$ by hypothesis, we have

$$I''(x) = f(x) + \int_0^x \psi''(x-t)\,f(t)\,dt.$$

Substitution of the expressions for $I(x)$, $I'(x)$, $I''(x)$ in the left-hand side of equation (1) gives

$$f(x) + \int_0^x \psi''(x-t)\,f(t)\,dt + a_1 \int_0^x \psi'(x-t)\,f(t)\,dt +$$

$$+ a_2 \int_0^x \psi(x-t)\,f(t)\,dt.$$

On bringing the constant coefficients a_1 and a_2 under the integral sign and combining all these integrals, we get

$$f(x) + \int_0^x [\psi''(x-t) + a_1\psi'(x-t) + a_2\psi(x-t)]\,f(t)\,dt.$$

But the integral term vanishes identically, since the expression in square brackets under the integral sign is identically zero. It is obtained by substituting $\psi(x-t)$ in the left-hand side of equation (1), whilst $\psi(x-t)$ is a solution of the corresponding homogeneous equation by hypothesis (the reader should not be worried by the fact that the argument of function ψ is $x-t$: if a homogeneous equation is satisfied by a function $\psi(x)$, it is evidently satisfied by $\psi(x-t)$).

Thus we have obtained $f(x)$ as a result of substituting $I(x)$ in the left-hand side of (1); this implies that $I(x)$ is a solution of (1). It is clear from the expression for $I(x)$ that $I(0) = 0$, $I'(0) = 0$.

We have thus shown that

$$y(x) = \int_0^x \psi(x-t)\,f(t)\,dt \qquad (2)$$

is a particular solution of equation (1).

Formula (2) holds for the linear equation

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_n y = f(x)$$

with constant coefficients and any order n if $\psi(x)$ is understood to be the solution of the corresponding homogeneous equation subject to the conditions

$$\psi(0) = \psi'(0) = \cdots = \psi^{(n-2)}(0) = 0, \quad \psi^{(n-1)}(0) = 1.$$

Formula (2) gives the particular solution $y(x)$ satisfying the conditions $y(0) = y'(0) = \cdots = y^{(n-1)}(0) = 0$.

It is worth pointing out that the integral in (2) is not affected by interchanging the arguments of functions f and ψ. For, let $x - t = \tau$; then

$$y = \int_0^x \psi(\tau) f(x - \tau)\, d\tau$$

or, on returning to the notation t for the variable of integration:

$$y = \int_0^x \psi(t) f(x - t)\, dt.$$

The two expressions for y give us

$$y = \frac{1}{2} \int_0^x [\psi(x - t) f(t) + \psi(t) f(x - t)]\, dt, \tag{3}$$

which sometimes proves to be more convenient to evaluate.

It may be mentioned that formula (2) can be deduced by using the method of variation of the arbitrary constants.

The use of formula (2) for seeking solutions has some advantage over the method of variation of the arbitrary constants in cases where we want to solve a number of linear non-homogeneous equations of the same type with different free terms. Apart from this, formula (2) is convenient from the theoretical point of view, since it gives explicitly the dependence of the solution of the non-homogeneous equation on its free term.

Example 1. An equation often encountered is

$$y'' + \omega^2 y = f(x).$$

The general solution of the corresponding homogeneous equation has the form (see Sec. 204, III):

$$y = C_1 \cos \omega x + C_2 \sin \omega x.$$

We find from this that

$$\psi(t) = \frac{1}{\omega} \sin \omega t$$

i.e. $$y = \frac{1}{\omega} \int_0^x \sin \omega(x - t) f(t)\, dt = \frac{1}{\omega} \int_0^x \sin \omega t\, f(x - t)\, dt$$

is a particular solution of the given equation.

Example 2. We take the equation discussed at the end of Sec. 205, II:

$$y'' + y = 4x \sin x.$$

We easily find from the general solution $y = C_1 \cos x + C_2 \sin x$ of the corresponding homogeneous equation that $\psi(x) = \sin x$.

We make use of formula (3):

$$y = 2 \int_0^x [t \sin t \sin (x - t) + (x - t) \sin t \sin (x - t)]\, dt$$
$$= 2x \int_0^x \sin t \sin (x - t)\, dt.$$

We obtain on transforming the integrand:

$$y = x \int_0^x [\cos (2t - x) - \cos x]\, dt$$
$$= x \left[\int_0^x \cos (2t - x)\, dt - \cos x \int_0^x dt \right],$$

i.e. $$y = x(\sin x - x \cos x),$$

which is what we found previously.

Suppose now that we have an equation with the same left-hand side but with $\tan x$ on the right-hand side. The special methods described in Sec. 204, I and II, cannot be used here. Formula (2) gives

$$y = \int_0^x \sin (x - t) \tan t\, dt = \int_0^x (\sin x \cos t - \cos x \sin t) \tan t\, dt$$
$$= \sin x \int_0^x \sin t\, dt - \cos x \int_0^x \frac{\sin^2 t}{\cos t}\, dt$$
$$= \sin x\, (1 - \cos x) - \cos x \ln \frac{1 + \sin x}{\cos x} + \sin x \cos x,$$

i.e. $$y = \sin x + \cos x \ln \frac{\cos x}{1 + \sin x}.$$

207. Vibrations. Resonance.

I. MECHANICAL VIBRATIONS. Vibrational problems are of considerable importance in present-day science and engineering. In the majority of cases vibratory phenomena are described by second order linear differential equations (linear vibrations), which have *constant* coefficients in the simplest conditions. We shall start by discussing mechanical vibrations.

Suppose a force acts on a moving body or system of bodies (we shall take the mass as unity for simplicity) in such a way as to tend to return the system to its equilibrium state, the magnitude of the force being proportional to the deviation from the equilibrium position. We write s for the distance from the equilibrium position at the instant t, and ω^2 for the coefficient of proportionality; then the magnitude of the force is $\omega^2 s$. A force of this kind is generally excited inside the actual system and is known as the *restoring force*; ω is called the *coefficient of restoration*.

We suppose further that resistance is present in the motion (resistive media, joints etc.). Let the force due to this be proportional to the velocity of the motion: its direction is opposite to that of the motion and its magnitude is equal to $2ks'$, where $2k$ denotes the coefficient of proportionality. This force is called the *resistive force* and k is the *coefficient of resistance*.

Finally, let an external force, given as a function $f(t)$ of time t, act on the system. It is termed the disturbing force (more precisely, the external disturbing force).

We form the differential equation of motion. When finding the force equilibrium we evidently need to take the restoring and resistive forces with the minus sign, since the first is directed opposite to the direction for measuring s, whilst the second is opposed to the velocity. We therefore arrive at the equation

$$s'' = -2ks' - \omega^2 s + f(t)$$

or

$$s'' + 2ks' + \omega^2 s = f(t). \qquad (*)$$

Mechanical vibrations of the above type are thus described by second order linear differential equations with constant coefficients.

If there is no external disturbing force ($f(t) \equiv 0$), the vibration is said to be free; equation (*) becomes homogeneous for a free vibration. When disturbing forces are present the vibration is said to be forced; this type of vibration is described by non-homogeneous equation (*), the right-hand side $f(t)$ of which is sometimes called the disturbing term.

The restoring force may be an elastic force. This case is obtained in practice say in the following circumstances. A load is suspended from a vertical spring and is exactly balanced by the elastic force of the spring. If the load is displaced from the equilibrium position by a vertical force it will then vibrate, the vibration being free if there are no auxiliary forces acting on the load or complete system, and forced if such forces are present (say when the point of suspension of the spring is displaced vertically according to some law). The spring reaction (restoring force) will, by Hooke's law, be proportional to the displacement (deformation). The restoration coefficient is the "stiffness" of the spring — a measure of the strength of the elastic connection. We considered free elastic vibration without resistance in Sec. 200, III, where it was shown to be a simple harmonic motion. We investigate the problem in more general terms below.

II. INVESTIGATION OF FREE VIBRATIONS. We shall apply our knowledge of the solutions of linear equations with constant coefficients to the study of vibrations. Let us start with the free vibrations

$$s'' + 2ks' + \omega^2 s = 0.$$

The roots of the characteristic equation are $-k \pm \sqrt{k^2 - \omega^2}$. Three cases are possible:

(1) *The resistance coefficient is greater than the restoration coefficient:* $k > \omega$. The solution here will be

$$s = C_1 e^{-\delta_1 t} + C_2 e^{-\delta_2 t},$$

where $\delta_1 = k - \sqrt{k^2 - \omega^2}$, $\delta_2 = k + \sqrt{k^2 - \omega^2}$.

We see that no *vibration* occurs; as t increases the displacement s diminishes and tends to zero as $t \to \infty$; the system tends to equilibrium monotonically (in practice it is reached at a finite instant). We are concerned with a non-periodic *damped* vibration. This case is evidently explained by the fact that the effect of the resistance force (braking the motion) so far exceeds the effect of the restoring force (producing vibration) that the motion is damped out before the system passes the equilibrium position.

(2) *The resistance coefficient is equal to the restoration coefficient:* $k = \omega$. The solution is now

$$s = e^{-kt}(C_1 + C_2 t).$$

This case is the same as the previous one, in the sense that s is a decreasing function as from a certain t, and $s \to 0$ as $t \to \infty$.

(3) *The coefficient of resistance is less than the restoration coefficient:* $k < \omega$. The solution is in this case

$$s = e^{-kt}(C_1 \cos k_1 t + C_2 \sin k_1 t), \quad \text{where} \quad k_1 = \sqrt{\omega^2 - k^2},$$

or

$$s = A e^{-kt} \sin(k_1 t + \varphi_0),$$

where A and φ_0 are arbitrary constants.

The system now in fact vibrates, the vibrations being termed damped harmonic. In contrast with purely harmonic vibrations, the vibration amplitude $A e^{-kt}$ is not a constant but depends on time; it tends to zero at $t \to \infty$. The system thus tends to equilibrium, though not monotonically — as oscillates about the equilibrium position with gradually diminishing amplitude (exponential damping). The logarithm of the amplitude $\ln A - kt$ decreases with constant velocity equal to k and is known as the *logarithmic decrement* of the damped harmonic vibration. As in the case of harmonic vibrations, $k_1 = \sqrt{\omega^2 - k^2}$ is termed the frequency, and φ_0 the initial phase of the damped harmonic vibration.

We shall not make a detailed investigation since this was done in Sec. 68 (for $A = 1$, $k = 1$, $k_1 = 1$, $\varphi_0 = 0$).

As regards the physical picture, vibrations are evidently excited because the restoration force exceeds the resistive force. But the vibrations are damped, since resistance is always present. If we assume no resistance at all ($k = 0$), we arrive at the familiar case of non-damped harmonic vibration (with a sine law).

III. INVESTIGATION OF FORCED VIBRATIONS. RESONANCE. We turn to the forced vibrations

$$s'' + 2ks' + \omega^2 s = f(t).$$

The nature of the motion naturally depends a great deal on the nature of the impressed external force. It is easy to imagine a disturbing force which will predominate decisively over the forces of restoration and resistance and will excite any desired motion. We shall only consider here a case of practical interest, when the disturbing force is periodic (sinusoidal): $f(t) = p \sin \omega_1 t$, and the resistance coefficient is less than the restoration coefficient: $k < \omega$.

If we do not take into account the "internal" forces of the system, i.e. the resistance and restoration forces, it is clear that the system will perform simple harmonic vibrations of the frequency ω_1 when under the influence of a single "external" sinusoidal force $p \sin \omega_1 t$. Whereas if we neglect the external disturbing force, the system will perform, as we have seen, damped ($k \neq 0$) or undamped ($k = 0$) harmonic vibrations with frequency $k_1 = \sqrt{\omega^2 - k^2}$. These free vibrations are known as the proper vibrations of the system, whilst $k_1 = \sqrt{\omega^2 - k^2}$ is the proper frequency as distinct from the "disturbing" frequency ω_1.

We now consider the motion of the system when internal and external forces act simultaneously.

We note first of all that, by the properties of linear non-homogeneous equations, any forced vibration is the result of finding the proper vibration of the system and any previously chosen forced vibration (Sec. 203, THEOREM 1). Let us find a particular forced vibration, i.e. a particular solution of the non-homogeneous equation

$$s'' + 2ks' + \omega^2 s = p \sin \omega_1 t.$$

Let $k \neq 0$; then $\omega_1 i$ is not a root of the characteristic equation and the solution can be sought in the form $a \cos \omega_1 t + b \sin \omega_1 t$. We find coefficients a and b by the familiar method and arrive at the solution

$$s = A_1 \sin(\omega_1 t + \varphi_1),$$

where

$$A_1 = \sqrt{a^2 + b^2} = \frac{p}{\sqrt{4k^2\omega_1^2 + (\omega^2 - \omega_1^2)^2}} \text{ and } \tan \varphi_1 = \frac{a}{b} = \frac{-2k\omega_1}{\omega^2 - \omega_1^2}.$$

Thus *one of the forced vibrations is a simple harmonic vibration of amplitude* A_1, *frequency* ω_1 *and initial phase* φ_1. Every other is obtained by imposing this vibration on the proper vibration of the system:

$$s = A e^{-kt} \sin(k_1 t + \varphi_0) + A_1 \sin(\omega_1 t + \varphi_1).$$

In view of the fact that the first term tends to zero fairly rapidly as t increases, the motion will be described in essence after a certain time only by the second term. This fact is bound up with an interesting phenomenon called resonance that may arise with forced vibrations.

We consider the amplitude of the forced vibration A_1 and suppose that the proper frequency of the system ω is fixed. The expression for A_1 may be written as

$$A_1 = \frac{p}{\omega^2 \sqrt{\frac{4k^2}{\omega^2}\left(\frac{\omega_1}{\omega}\right)^2 + \left[1 - \left(\frac{\omega_1}{\omega}\right)^2\right]^2}} = \frac{p}{\omega^2 \sqrt{\alpha^2 q^2 + (1 - q^2)^2}},$$

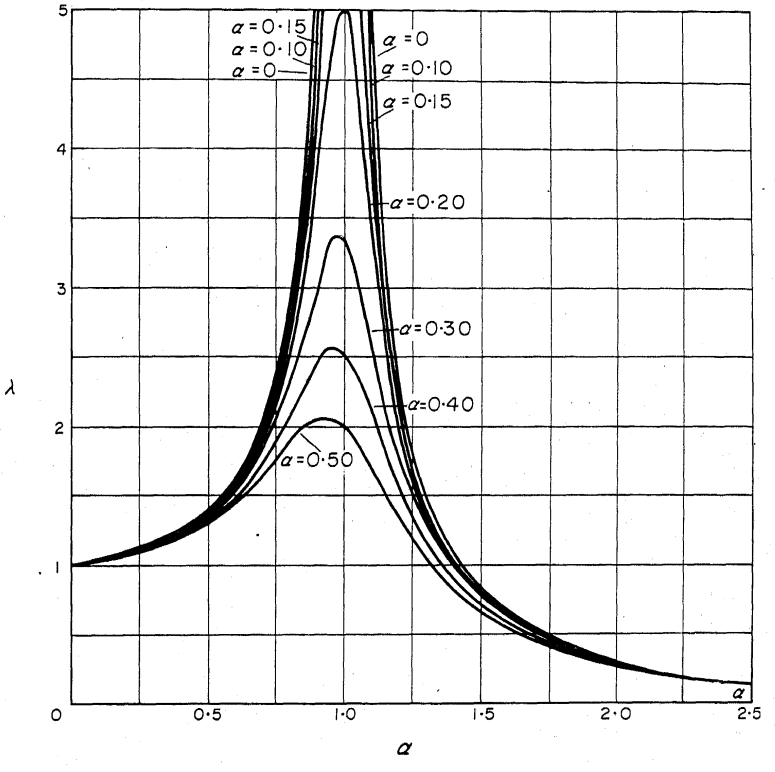

FIG. 75

where $\alpha = 2k/\omega$, $q = \omega_1/\omega$*. The amplitude A_1 is proportional to

$$\lambda = \lambda(q) = \frac{1}{\sqrt{\alpha^2 q^2 + (1 - q^2)^2}}.$$

The nature of the variation of λ with q may be seen from graphs of function $\lambda(q)$ corresponding to various values of α. Figure 75 shows the graphs of $\lambda = \lambda(q)$ for $\alpha = 0{\cdot}50$; $0{\cdot}40$; $0{\cdot}30$; $0{\cdot}20$; $0{\cdot}15$; $0{\cdot}10$; 0. As a rule α is fairly

* Since the frequency is the reciprocal of the period, we can also say that q is equal to the ratio T/T_1 of the proper period of the system T to the "forced period" T_1.

small, since the resistance coefficient k is usually only a small fraction of the restoration coefficient ω.

Investigation shows that, as q increases, i.e. as the disturbing frequency ω_1 increases, λ and the amplitude of the stationary forced vibration increase rapidly to a maximum then diminish rapidly, tending to zero as $q \to \infty$ $(\omega_1 \to \infty)$. This maximum can be fairly large with small values of the resistance coefficient k. When the disturbing frequency is close to the proper frequency of the system (and all the more when they are the same) and when the resistance coefficient is relatively small, the system will perform (after a certain interval of time) harmonic vibrations of very large amplitude, which can be considerably greater than the amplitude of the sinusoidal force exciting the actual vibration of the system. This phenomenon of a rapid rise in the vibration amplitude under the action of even quite a small external disturbance is known as *resonance*. The curves of Fig. 75 are called *resonance curves*.

Resonance plays an important role in engineering and physics. Any elastic body (for instance, any structure) has its own definite proper frequency of vibration, depending only on the properties of the body; let us imagine that the body is shifted from its equilibrium position by the action of an external force; if this force is sinusoidal (or in general periodic), and its frequency is close to the proper frequency of the body, the effect of the disturbing force on the body may prove substantial and destructive, no matter how small it is. When designing various structures (machines, bridges, ships, aeroplanes etc.) special attention is paid to considerations of strength relevant to resonance. Resonance accounts for a phenomenon which is often seen in practice, when an elastic body (say a bridge) is caused to break up due to a fairly small "sway".

In the absence of resistance, i.e. when $k = 0$ (though this is impossible in practice), the amplitude of the forced vibration is equal to $p/|\omega^2 - \omega_1^2|$ and becomes infinite when the disturbing and proper frequencies coincide $(\omega_1 = \omega)$ (see Fig. 75, curve $\alpha = 0$). In the case when $\omega_1 \neq \omega$, the solution of the differential equation of the vibration $s'' + \omega^2 s = p \sin \omega_1 t$ is got from the general solution found above by assigning the value 0 to k. But if $\omega_1 = \omega$, the general solution does not apply, since ωi now becomes a root of the characteristic equation $r^2 + \omega^2 = 0$. The reader may easily show the solution to be

$$s = A \sin(\omega t + \varphi_0) - \frac{p}{2\omega} t \cos \omega t.$$

The second term (called the "secular term") shows that the vibration amplitude increases indefinitely in the course of time, and shows that resonance is present.

IV. Oscillations in electrical circuits. A picture fully analogous to that of mechanical vibrations is given by the electric current oscillations in an open circuit of resistance R, inductance L and capacity C with an externally impressed voltage $v = v(t)$. A similar problem has already been considered in Sec. 194, where, however, no account was taken of capacity. In the general case part of the total voltage is due to the capacity. It is shown in physics that this part is given by $\frac{1}{C} \int i \, dt$, where i is the current.

We thus arrive at the equation

$$Li' + Ri + \frac{1}{C}\int i\,dt = v(t),$$

whence we obtain by differentiation with respect to t a second order differential equation with constant coefficients:

$$Li'' + Ri' + \frac{1}{C}i = v'(t).$$

This differential equation for the current flow is analogous to the equation describing mechanical vibrations.

If the external electromotive force is constant (or zero), the equation will be homogeneous. The corresponding current is necessarily damped since resistance is present ($R \neq 0$). It is described as transient in electrical engineering. On extending further the analogy with mechanical vibrations, we can say that this current (assuming that $R/2L < \sqrt{1/CL}$) gives the "proper" oscillations of the circuit. If an external disturbing voltage is present, such that $v'(t)$ is not identically zero, the non-homogeneous equation has a particular solution $i = i(t)$, which gives the ("forced") current, which is described as stationary. Finally, the general solution of the equation, i.e. the dependence of the current on time with any conditions, is obtained by adding the "transient" and "stationary" currents. The current given by this solution is called the total current. It is quite clear that, after a certain length of time, the "transient" current will have no practical effect on the flow of electricity, and the total current becomes the same as the "stationary."

Assuming that the voltage $v(t)$ impressed on the circuit is a sinusoidal function of time, we discover as in the case of mechanical vibrations that the "stationary" current is also a sinusoidal function of time. It gives the "forced" oscillations, whose amplitude depends in essence on the difference between the frequency of the "disturbing" voltage and the frequency of the "transient" damped current. We also encounter resonance here. There is no need to go into further details, since they are similar from the mathematical point of view to those described for mechanical vibrations, and the special problems of electrical engineering have no place in this book.

6. Supplementary Problems

208. Some Linear Equations Leading to Equations with Constant Coefficients. We shall mention some simple examples of linear equations with variable coefficients, the integration of which reduces to solution of linear equations with constant coefficients.

I. We take the second order linear equation

$$x^2y'' + a_1xy' + a_2y = f(x),$$

where a_1 and a_2 are constant. It is known as *Euler's equation.*

We consider it for $x > 0$. Replacing the independent variable in accordance with $x = e^t$ or $t = \ln x$ transforms Euler's equation

to another second order linear equation, but with constant coefficients. For we have:

$$y' = \frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx} = \frac{dy}{dt}\frac{1}{x},$$

$$y'' = \frac{dy'}{dx} = \frac{dy'}{dt}\frac{dt}{dx} = \left(\frac{d^2y}{dt^2}\frac{1}{x} - \frac{dy}{dt}\frac{1}{x^2}\frac{dx}{dt}\right)\frac{dt}{dx}$$

$$= \frac{d^2y}{dt^2}\frac{1}{x^2} - \frac{dy}{dt}\frac{1}{x^2}.$$

Substitution in the equation gives

$$\frac{d^2y}{dt^2} - \frac{dy}{dt} + a_1\frac{dy}{dt} + a_2y = f(e^t),$$

i.e.
$$y'' + (a_1 - 1)y' + a_2y = f(e^t),$$

where the derivatives of y are taken with respect to the new variable t. Having found y as a function of t from the equation, we obtain the required solution of Euler's equation by replacing t by $\ln x$.

Example. $x^2y'' - 2y = 2x\ln x$.

The substitution $x = e^t$, $t = \ln x$ gives:

$$y'' - y' - 2y = 2te^t.$$

It is easy to find a particular solution of the non-homogeneous equation:

$$y = -\left(t + \frac{1}{2}\right)e^t;$$

the general solution of the corresponding homogeneous equation will be $y = C_1e^{-t} + C_2e^{2t}$. Hence the general solution of the transformed equation with constant coefficients is

$$y = C_1e^{-t} + C_2e^{2t} - \left(t + \frac{1}{2}\right)e^t.$$

On returning to the variable x, we get the general solution of the given Euler's equation:

$$y = C_1\frac{1}{x} + C_2x^2 - \left(\ln x + \frac{1}{2}\right)x.$$

The same substitution $x = e^t$ reduces the n-th order linear Euler equation

$$x^n y^{(n)} + a_1 x^{n-1} y^{(n-1)} + \cdots + a_{n-1} x y' + a_n y = f(x),$$

where $a_1, \ldots, a_{n-1}, a_n$ are constants, to a linear n-th order equation with constant coefficients.

II. BESSEL'S EQUATION

$$x^2 y'' + x y' + (x^2 - n^2) y = 0, \quad n = \text{const},$$

can be reduced in the case when $n = \frac{1}{2}$ to a linear equation with constant coefficients. We substitute for the required function y in accordance with

$$y = \frac{z}{\sqrt{x}} = z x^{-\frac{1}{2}}.$$

We have:

$$y' = z' x^{-\frac{1}{2}} - \frac{1}{2} z x^{-\frac{3}{2}},$$

$$y'' = z'' x^{-\frac{1}{2}} - z' x^{-\frac{3}{2}} + \frac{3}{4} z x^{-\frac{5}{2}}.$$

Substitution in the equation gives

$$z'' x^{\frac{3}{2}} - z' x^{\frac{1}{2}} + \frac{3}{4} z x^{-\frac{1}{2}} + z' x^{\frac{1}{2}} - \frac{1}{2} z x^{-\frac{1}{2}} + \left(x^2 - \frac{1}{4}\right) z x^{-\frac{1}{2}} = 0,$$

i.e.

$$z'' + z = 0.$$

This equation has the fundamental system of solutions: $z_1 = \cos x$, $z_2 = \sin x$. Hence the fundamental system of solutions for Bessel's equation with $n = \frac{1}{2}$ is

$$y_1 = \frac{\cos x}{\sqrt{x}}, \quad y_2 = \frac{\sin x}{\sqrt{x}}.$$

This fact should already be familiar from Sec. 202.

209. Systems of Differential Equations.

Definition. **A *system of differential equations* is a set of equations containing the same independent variables, required functions and derivatives.**

A system of two first order ordinary linear equations (i.e. with one independent variable) with respect to two required functions y and z has the form

$$a_0 y' + b_0 z' + a_1 y + b_1 z = f_1,$$
$$\alpha_0 y' + \beta_0 z' + \alpha_1 y + \beta_1 z = f_2,$$

where a_0, b_0, a_1, b_1, α_0, β_0, α_1, β_1, f_1, f_2 are given functions of the independent variable x.

The problem consists in finding functions $y = y(x)$, $z = z(x)$ satisfying both equations.

We can find from the two equations say z as a function of x, y and y'. On next differentiating the expression obtained and substituting for z and z' in either equation, we arrive at a linear differential equation which is in general of the second order in y. Having found y from it, we get the second required function z without quadratures. Thus:

The general solution of a system of first order linear equations depends on two arbitrary constants.

Example.

$$y' + z' - \frac{2}{x^2}y - z = x,$$

$$xy' - z' + \frac{2}{x^2}y - xz = x^2.$$

Addition gives

$$(1 + x)y' - (1 + x)z = x(1 + x),$$

whence $$z = y' - x \quad \text{and} \quad z' = y'' - 1.$$

Substitution in the first equation of the system leads to the Euler equation

$$x^2y'' - 2y = x^2.$$

Integration of this (see Sec. 208) gives

$$y = \frac{x^2}{3}\ln x + \frac{C_1}{x} + C_2x^2$$

i.e. $$z = y' - x = \frac{2}{3}x(\ln x - 1) - \frac{C_1}{x^2} + 2C_2x.$$

If the given equations have constant coefficients, the equation at which we arrive, of the second order in one of the required functions, will also have constant coefficients.

The above method can also be applied to the integration of a system of two non-linear differential equations of the first order with two unknown functions.

Let the system be

$$F_1(x, y, y', z, z') = 0, \quad F_2(x, y, y', z, z') = 0.$$

We find z as a function of x, y, y' from this system:

$$z = \varphi(x, y, y'),$$

and hence find z' as a function of x, y, y', y'' by differentiation. On introducing the expressions found for z and z' into one equation of the system, we obtain a second order equation in the required function y. Knowing this, the other function z is found without quadratures.

In certain cases all these operations can be simplified by virtue of individual features of the system. For instance, if the system is

$$y' = \frac{x}{yz}, \qquad z' = \frac{x}{y^2},$$

we obtain on dividing the second equation by the first:

$$\frac{z'}{y'} = \frac{z}{y}, \qquad \text{i.e.} \qquad \frac{dz}{z} = \frac{dy}{y},$$

whence $z = C_1 y$. On substituting this say in the first equation of the system, we again get an equation with separated variables:

$$y^2 dy = \frac{x}{C_1} dx.$$

Hence

$$y = \sqrt[3]{\frac{3}{2C_1} x^2 + C_2} \quad \text{and} \quad z = C_1 \sqrt[3]{\frac{3}{2C_1} x^2 + C_2}.$$

We have not had to solve an equation of order higher than the first in this example. In the case when the equations of the system are of higher order than the first, their integration in general leads to solving equations of higher order than the second.

We shall consider an electrical problem leading to a system of two linear second order equations with constant coefficients.

Let two circuits be given, the resistance, inductance and capacity of which will be denoted by R_1, L_1, C_1 and R_2, L_2, C_2 respectively. Suppose further that the circuits are electrically connected: each induces an electromotive force into the other. If we write M for the constant coefficient of mutual inductance, the voltage induced into the first circuit will be $M\, di_2/dt$, and into the second $M\, di_1/dt$, where $i_1 = i_1(t)$, $i_2 = i_2(t)$ denote the currents in the first and second circuits respectively.

Let there be no external electromotive force in either circuit. The current flow in the two circuits will in this case be determined by the following differential equations (see Sec. 207, IV):

$$L_1 i_1' + R_1 i_1 + \frac{1}{C_1}\int i_1\,dt + M i_2' = 0,$$

$$L_2 i_2' + R_2 i_2 + \frac{1}{C_2}\int i_2\,dt + M i_1' = 0_2,$$

or after differentiation with respect to t:

$$L_1 i_1'' + R_1 i_1' + \frac{1}{C_1} i_1 + M i_2'' = 0,$$

$$L_2 i_2'' + R_2 i_2' + \frac{1}{C_2} i_2 + M i_1'' = 0.$$

We solve this system by proceeding as in the case of a system of first order equations. We use elimination and differentiation to obtain an equation containing only one of the required functions and its derivatives. In fact, we find i_2'' from the second equation and substitute in the first:

$$(L_1 L_2 - M^2)\, i_1'' + L_2 R_1 i_1' + \frac{L_2}{C_1} i_1 - M R_2 i_2' - \frac{M}{C_2} i_2 = 0.$$

We now differentiate with respect to t and replace $M i_2''$ by its expression found from the first equation of the system; this gives

$$(L_1 L_2 - M^2)\, i_1''' + (L_1 R_2 + L_2 R_1)\, i_1'' + \left(\frac{L_2}{C_1} + R_1 R_2\right) i_1' + \\ + \frac{R_2}{C_1} i_1 - \frac{M}{C_2} i_2' = 0.$$

Finally, renewed differentiation and replacement of $M i_1''$ leads to a linear fourth order equation with constant coefficients in the required function $i_1(t)$:

$$(L_1 L_2 - M^2)\, i_1^{(\mathrm{IV})} + (L_1 R_2 + L_2 R_1)\, i_1''' + \left(\frac{L_1}{C_2} + \frac{L_2}{C_1} + R_1 R_2\right) i_1'' + \\ + \left(\frac{R_1}{C_2} + \frac{R_2}{C_1}\right) i_1' + \frac{1}{C_1 C_2} i_1 = 0.$$

Having found i_1 from this, function i_2 is obtained from quite a simple second order equation or even without quadratures at all if use is made of the intermediate equations. For we can eliminate i_2' from the two equations obtained in the process of elimination and differentiation, then find directly an expression for i_2 in terms of i_1 and its derivatives.

If the capacities are neglected, the current in two circuits linked by mutual inductance will be described by a system of first order linear equations.

We remark in conclusion that a system of three, four, etc., equations with the corresponding number of required functions is usually solved in practice by essentially the same method as for a system of two equations. The system is reduced with the aid of successive eliminations and differentiations to a system of equations, each of which contains in the last analysis only one of the required functions and its derivatives.

The orders of the equations can be considerably increased during this process. It is not possible to give a logical scheme in advance for these reductions. It depends on the nature of the system and its special features.

CHAPTER XV

TRIGONOMETRIC SERIES

1. Trigonometric Polynomials

210. The Problem. A series is trigonometric if it has the form

$$a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \cdots$$

$$\cdots + (a_n \cos nx + b_n \sin nx) + \cdots = \sum_{n=0}^{\infty} (a_n \cos nx + b_n \sin nx).$$

Various problems of pure and applied mathematics lead to such infinite series. Trigonometric series compose the second most important class of functional series (the first is the class of power series, see Chapter IX).

We have already encountered problems in which trigonometric functions of multiple arguments have to be added. Thus if x denotes time, the function $s = A \sin(\omega x + x_0)$ describes the familiar *simple harmonic vibration with frequency** ω. The function $A \sin(\omega x + x_0)$ (and its graph) is called a simple harmonic. When several forces, each of which produces a harmonic vibration, act simultaneously, the resultant vibration is got by adding the simple harmonics. For instance, we obtain in the case of two forces (see Sec. 25):

$$s = A_1 \sin(\omega_1 x + x_1) + A_2 \sin(\omega_2 x + x_2).$$

This function (and its graph) is called a *compound harmonic.*

Let the frequencies ω_1 and ω_2 of the superimposed vibrations be equal, $\omega_1 = \omega_2$. The resultant motion is now another simple harmonic vibration, and the "compound" harmonic remains simple.

If frequencies ω_1 and ω_2 are not commensurable, the resultant motion is not even periodic (let alone harmonic); whereas if they are commensurable: $\omega_1 = r_1\omega$, $\omega_2 = r_2\omega$, where r_1 and r_2 are in-

* This is the "circular" frequency, i.e. the number of periods of the function (vibration) contained in 2π units of the independent variable (time).

tegers, superimposition of the two vibrations leads to a motion which, though not simple harmonic, is periodic with period $2\pi/\omega$. For in this case the compound harmonic

$$A_1 \sin(r_1\omega x + x_1) + A_2 \sin(r_2\omega x + x_2)$$

is a number which does not change its value on adding $2\pi/\omega$ to any given value of x; since $2\pi/\omega$ is the least positive number possessing this property, it is the period. There will be no loss of generality if we assume $\omega = 1$ (we can always reduce to this by replacing the variable in accordance with $x' = \omega x$).

Let us consider the periodic function

$$y = A_1 \sin(r_1 x + x_1) + A_2 \sin(r_2 x + x_2)$$

(r_1, r_2 are integers) with period equal to 2π. We saw in Chapter I that such a function can be very much *sui generis* and widely different from the simple harmonics. It was observed that a sum of simple harmonics

$$y = A_1 \sin(r_1 x + x_1) + A_2 \sin(r_2 x + x_2) + \cdots$$
$$\cdots + A_n \sin(r_n x + x_n)$$

leads in general to quite different types of periodic function with different values of parameters A_k, x_k and integers r_k and n. This means in terms of mechanics that the superimposition of simple harmonic vibrations produces widely different periodic motions which have nothing in common with the simple harmonic vibrations.

The converse problem arose naturally: would it not be possible to choose simple harmonic vibrations in such a way that superimposition of them gave a previously assigned periodic motion, i.e. would it not be possible to represent any periodic motion as a "compound harmonic vibration". The mathematical problem consists in the possibility of representing a previously assigned periodic function as a "compound harmonic" by a suitable choice of the parameters of the simple harmonics.

One of the greatest achievements of nineteenth century mathematical physics was to provide a strict basis for an affirmative answer to this question. It was found that, provided an infinity of simple harmonics was brought in, i.e. use was made of an infinite series rather than a finite sum, any function with period say 2π be-

longing to a very wide class of functions could be analyzed into simple harmonics, i.e. could be written as the infinite series

$$f(x) = \sum_{n=0}^{\infty} A_n \sin(nx + x_n)$$

$$= A_0 \sin x_0 + A_1 \sin(x + x_1) + A_2 \sin(2x + x_2) + \cdots$$

$$\cdots + A_n \sin(nx + x_n) + \cdots$$

Each simple harmonic may be written as

$$A_n \sin(nx + x_n) = A_n \sin x_n \cos nx + A_n \cos x_n \sin nx$$

$$= a_n \cos nx + b_n \sin nx,$$

where we use the notation $a_n = A_n \sin x_n$, $b_n = A_n \cos x_n$. The infinite series of simple harmonics now becomes the trigonometric series

$$f(x) = a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \cdots$$

$$\cdots + (a_n \cos nx + b_n \sin nx) + \cdots$$

$$= \sum_{n=0}^{\infty} (a_n \cos nx + b_n \sin nx), \qquad (*)$$

where a_n, b_n are constants.

We shall indicate in Sec. 212 the conditions in which a given function can be written as the sum of an infinite convergent trigonometric series (*). When n is sufficiently large, the sum of the first n terms represents $f(x)$ to any desired degree of accuracy. The finite sum

$$s_n = \sum_{k=0}^{n} (a_k \cos kx + b_k \sin kx)$$

is called a *trigonometric polynomial of order n* (we are assuming that the last coefficients a_n and b_n do not vanish simultaneously: $a_n^2 + b_n^2 \neq 0$).

It should be noted that, since $f(x)$ (and series (*)) has period 2π, it is sufficient to investigate the convergence of series (*) to $f(x)$ in any interval of length 2π, say in $[-\pi, \pi]$; in any other pieces of the Ox axis function $f(x)$ (and the series) will strictly repeat its values and behaviour in the base interval $[-\pi, \pi]$.

211. Fourier Coefficients and their Properties.

I. FOURIER COEFFICIENTS. Let $f(x)$ be any trigonometric polynomial of order n:

$$f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx);$$

and let us find expressions for coefficients $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ in terms of function $f(x)$. (The free coefficient is written as $\frac{1}{2}a_0$ instead of a_0 since by this means, as we shall soon see, a single unified formula is obtained for all the coefficients.)

We shall require the following relationships (k and p are non-negative integers):

$$\int_{-\pi}^{\pi} \cos kx \cos px\, dx = \begin{cases} 0, & \text{if } k \neq p, \\ \pi, & \text{if } k = p \neq 0; \end{cases} \tag{A_1}$$

$$\int_{-\pi}^{\pi} \sin kx \sin px\, dx = \begin{cases} 0, & \text{if } k \neq p, \\ \pi, & \text{if } k = p \neq 0; \end{cases} \tag{A_2}$$

$$\int_{-\pi}^{\pi} \cos kx \sin px\, dx = 0 \text{ (any } k \text{ and } p) \tag{A_3}$$

Let us prove relationship (A_1).

We have by a familiar trigonometric formula:

$$\cos kx \cos px = \frac{1}{2}[\cos(k+p)x + \cos(k-p)x]. \tag{1}$$

If $k \neq p$, we have

$$\int_{-\pi}^{\pi} \cos kx \cos px\, dx$$

$$= \frac{1}{2}\int_{-\pi}^{\pi} \cos(k+p)x\, dx + \frac{1}{2}\int_{-\pi}^{\pi} \cos(k-p)x\, dx$$

$$= \frac{1}{2}\frac{\sin(k+p)x}{k+p}\Bigg|_{-\pi}^{\pi} + \frac{1}{2}\frac{\sin(k-p)x}{k-p}\Bigg|_{-\pi}^{\pi} = 0,$$

since $k + p$ and $k - p$ are integers. If $k = p$, we have

$$\int_{-\pi}^{\pi} \cos^2 px\, dx = \frac{1}{2}\int_{-\pi}^{\pi} (1 + \cos 2px)\, dx = \frac{1}{2} x \Big|_{-\pi}^{\pi} +$$

$$+ \frac{1}{2} \frac{\sin 2px}{2p} \Big|_{-\pi}^{\pi} = \pi.$$

Relationships (A_1) are thus proved. Formulae (A_2) and (A_3) are proved in a similar way; the necessary working may be carried out by the reader.

To find coefficient a_p, where p is any positive integer, we multiply both sides of equation (1) by $\cos px$ and take the integral from $-\pi$ to π:

$$\int_{-\pi}^{\pi} f(x) \cos px\, dx = \frac{a_0}{2}\int_{-\pi}^{\pi} \cos px\, dx +$$

$$+ \sum_{k=1}^{n} \left(a_k \int_{-\pi}^{\pi} \cos kx \cos px\, dx + b_k \int_{-\pi}^{\pi} \sin kx \cos px\, dx \right).$$

In view of formulae (A_1) and (A_3), all the integrals on the right-hand side vanish, except for the integral with coefficient a_p, this being equal to π. Thus

$$\int_{-\pi}^{\pi} f(x) \cos px\, dx = \pi a_p .$$

This gives for a_p:

$$a_p = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \cos px\, dx. \tag{2}$$

If $p = 0$, we have

$$\int_{-\pi}^{\pi} f(x)\, dx = \frac{a_0}{2}\int_{-\pi}^{\pi} dx = \pi a_0 ,$$

and we obtain an expression which is the particular case when $p = 0$ of the general formula for coefficient a_p:

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, dx .$$

To find b_p, we multiply both sides of equation (1) by $\sin px$ and again integrate from $-\pi$ to π:

$$\int_{-\pi}^{\pi} f(x) \sin px \, dx$$

$$= \frac{a_0}{2} \int_{-\pi}^{\pi} \sin px \, dx + \sum_{k=1}^{n} \left(a_k \int_{-\pi}^{\pi} \cos kx \sin px \, dx + b_k \int_{-\pi}^{\pi} \sin kx \sin px \, dx \right).$$

In view of formulae (A_2) and (A_3) all the integrals on the right-hand side vanish except for the integral with coefficient b_p, which is equal to π. Thus

$$b_p = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin px \, dx . \tag{3}$$

Now suppose that $f(x)$ is not a trigonometric polynomial of order n. Equation (1) will no longer hold for $f(x)$; but we can form the n-th order trigonometric polynomial $\Phi_n(x)$, with coefficients defined in accordance with expressions (2) and (3) for $f(x)$. We assume here that $f(x)$ is an integrable* function in the interval $[-\pi, \pi]$.

We shall therefore have

$$\left.\begin{aligned} a_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx , \\ b_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx , \end{aligned}\right\} \quad k = 0, 1, 2, \ldots, n. \tag{E}$$

Definition. **The numbers a_k, b_k found in accordance with formulae (E) are called the *Fourier*** coefficients of order k of function $f(x)$, whilst the trigonometric polynomial $\Phi_n(x)$ with these coefficients:**

$$\Phi_n(x) = \tfrac{1}{2} a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

is the *n-th order Fourier polynomial of function $f(x)$*.

* This means that the integral of $f(x)$ exists in $[-\pi, \pi]$.

** J. B. Fourier (1768–1830) was a famous French mathematician and physicist, who applied trigonometric series systematically to the solution of problems of mathematical physics.

The above method of finding the coefficients was first given by L. P. EULER, whilst the extensive use of these coefficients for forming approximate trigonometric polynomials was due to Fourier.

We take the Fourier polynomial $\Phi_n(x)$ as an approximation to a given function $f(x)$, and write R_n for the error involved. Now,

$$f(x) = \Phi_n(x) + R_n. \tag{F}$$

This is the *Fourier formula*, R_n *being the remainder term.*

Let us compare Fourier's and Taylor's formulae. A necessary condition for the existence of the n-th order Taylor formula for $f(x)$ in an interval $[a, b]$ is that $f(x)$ have derivatives up to and including order $n + 1$ in this interval, whereas for the existence of the Fourier formula of any order of $f(x)$ in the interval $[-\pi, \pi]$ we only need to demand that $f(x)$ be integrable, since all the integrals occurring in the formula exist in this case. In particular, continuity of $f(x)$ obviously guarantees the existence of all the integrals. But we can easily widen the class of functions $f(x)$ for which Fourier's formula holds. In fact, nothing is changed in our arguments if $f(x)$ has a finite number of finite discontinuities in $[-\pi, \pi]$. In this case the integrands of all the integrals in Fourier's formula have the same properties and therefore exist.

However, the fact just mentioned can only be regarded as a really important and useful property of Fourier's formula provided it gives a good approximation, i.e. in other words, provided the remainder term can be made sufficiently small.

We shall see in the next article that this proviso is met by a very wide class of functions.

II. PROPERTY OF THE COEFFICIENTS.

THEOREM. **The Fourier coefficients of any function $f(x)$ which is integrable together with its square* tend to zero:**

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx\, dx = 0,$$

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx\, dx = 0.$$

* i.e. the integrals of both $f(x)$ itself and of its square $f^2(x)$ exist in the interval.

Proof. We consider the integral of the square of the remainder term of Fourier's formula:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} R_n^2\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi}[f(x) - \Phi_n(x)]^2 dx$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} f^2(x)\,dx - \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\Phi_n(x)\,dx + \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_n^2(x)\,dx.$$

We multiply the expression

$$\Phi_n(x) = \frac{1}{2}a_0 + \sum_{k=1}^{n}(a_k\cos kx + b_k\sin kx)$$

by $f(x)$ and integrate the product term by term:

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\Phi_n(x)\,dx = \frac{1}{2}a_0\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx +$$

$$+\sum_{k=1}^{n}\left(a_k\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx + b_k\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\,dx\right);$$

we obtain on taking formula (E) into account:

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\Phi_n(x)\,dx = \frac{1}{2}a_0^2 + \sum_{k=1}^{n}(a_k^2 + b_k^2).$$

We now square $\Phi_n(x)$ and again take the integral:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_n^2(x)\,dx = \frac{1}{4}a_0^2 +$$

$$+\sum_{k=1}^{n}\left(a_k^2\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos^2 kx\,dx + b_k^2\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin^2 kx\,dx\right) +$$

$$+\frac{1}{2\pi}2\int_{-\pi}^{\pi} P(x)\,dx,$$

where $P(x)$ denotes the sum of all possible paired products of terms in the expression for $\Phi_n(x)$. We conclude from expressions (A_1), (A_2) and (A_3) that the integral of each term in $P(x)$ is equal to zero, and each integral of $\cos^2 kx$ and $\sin^2 kx$ is equal to π. Consequently,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_n^2(x)\,dx = \frac{1}{2}\left[\frac{1}{2}a_0^2 + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right]$$

and therefore

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}R_n^2\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi}f^2(x)\,dx - \frac{1}{2}\left[\frac{1}{2}a_0^2 + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right].$$

Since the left-hand side of this equation is non-negative, the right-hand side must also be non-negative, i.e.

$$\frac{1}{2}a_0^2 + \sum_{k=1}^{n}(a_k^2 + b_k^2) \leqslant \frac{1}{\pi}\int_{-\pi}^{\pi}f^2(x)\,dx\,.$$

We observe that the right-hand side is independent of n, in other words the relationship holds for any n, so that the series obtained on the left when $n \to \infty$ is convergent. Thus

$$\frac{1}{2}a_0^2 + \sum_{k=1}^{\infty}(a_k^2 + b_k^2) \leqslant \frac{1}{\pi}\int_{-\pi}^{\pi}f^2(x)\,dx\,,$$

i.e. the sum of the series does not exceed the integral on the right-hand side. (We shall see below that the equals sign actually holds.) But since the series is convergent its general term $(a_n^2 + b_n^2)$ must tend to zero as $n \to \infty$, which is only possible if both $a_n \to 0$ and $b_n \to 0$. The theorem is proved.

It is worth noticing that equations (*) are interesting examples of passage to the limit in an integral when the limit sign cannot be taken under the integral sign: the integrands have no limits as $n \to \infty$ yet the integrals tend to zero.

Use will be made later of the properties of the coefficients given by the following theorem.

THEOREM. **The series with general terms**

$$\frac{|a_n|}{n} \quad \text{and} \quad \frac{|b_n|}{n}$$

are convergent,

Proof. The elementary inequality

$$AB \leqslant \frac{1}{2}(A^2 + B^2)$$

gives us, on setting $A = |a_n|$, $B = 1/n$:

$$\frac{|a_n|}{n} \leqslant \frac{1}{2}\left(a_n^2 + \frac{1}{n^2}\right).$$

Since the series $\sum_{n=1}^{\infty} a_n^2$ and $\sum_{n=1}^{\infty} 1/n^2$ are convergent (the first by virtue of the convergence proved above of $\sum_{1}^{\infty}(a_n^2 + b_n^2)$, the second by the familiar test (see Sec. 124)), the series $\sum_{n=1}^{\infty} |a_n|/n$ is also convergent. In the same way, it may be seen that the series $\sum_{n=1}^{\infty} |b_n|/n$ is convergent.

2. Fourier Series

212. Fundamental Theorems. We now turn to the fundamental propositions on the expansion of functions in trigonometric series.

Given the integrable function $f(x)$ in the interval $[-\pi, \pi]$, we shall find its Fourier coefficients of all orders and write in a purely formal way the trigonometric series with these coefficients.

Definition. **The series**

$$\tfrac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx),$$

where

$$\left.\begin{aligned} a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx, \\ b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx, \end{aligned}\right\} \quad n = 0, 1, 2, \ldots, \qquad \text{(E)}$$

s called the *Fourier series* of function $f(x)$.

We describe $f(x)$ as "generating" the series; this is written as

$$f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx),$$

the equals sign being avoided until the series on the right has been shown to be in fact *convergent to function* $f(x)$.

An important problem in the theory of trigonometric series relates to the conditions in which the Fourier series of $f(x)$ "represents" $f(x)$, i.e. is convergent to it. The direct way to solve this problem is obviously to investigate the remainder term R_n of the Fourier formula (Sec. 211, I):

$$f(x) = \Phi_n(x) + R_n.$$

Since $\Phi_n(x)$ is the n-th partial sum of the Fourier series, the fact that R_n tends to zero as $n \to \infty$ is equivalent to the convergence of the Fourier series to its generating function $f(x)$. The general theorems given below are in fact usually proved by investigating the remainder term R_n. We shall omit the actual proofs here*. Before giving the theorems we recall the following definitions (see Secs. 54 and 81):

A function $f(x)$ *is described as* smooth *in the interval* $[a, b]$ *if it is continuous along with its first derivative in the interval; the values of* $f'(x)$ *at the boundary points* $x = a$ *and* $x = b$ *are taken as the limits* $f'(a + 0)$ *and* $f'(b - 0)$ *to which* $f'(x)$ *tends when the independent variable tends to* a *and* b *from inside the interval.*

A function $f(x)$ *is said to be* piecewise smooth *in the interval* $[a, b]$ *if* $[a, b]$ *can be divided into a finite number of subintervals in each of which* $f(x)$ *is smooth.*

The graph of a smooth function is a continuous curve with a "*continuously rotating*" tangent; such a curve is described as smooth. The graph of a piecewise smooth function consists of a finite number of smooth arcs; such a curve is said to be piecewise smooth.

In view of these definitions a function which is piecewise smooth in interval $[a, b]$ can only have a finite number of singular points, at which either the function or its derivative has a discontinuity of the first kind.

We note that:

(1) if $f(x)$ is a piecewise smooth function, $f^2(x)$ is also piecewise smooth; (2) a function which is piecewise smooth in an interval is integrable in the interval.

THEOREM 1. **The Fourier series of a function $f(x)$, piecewise smooth in the interval $[-\pi, \pi]$ is convergent at every point x of the interval to $\frac{1}{2}[f(x-0) + f(x+0)]$ (i.e. in particular, of $f(x)$ itself at points**

* See this *Course*, 1950 and 1951 eds. See also Sec. 215 for a particular case.

where the function is continuous). At both ends to the interval the sum of the series is equal to the arithmetic mean of its limits as the independent variable tends to the end-values from inside the interval:

$$\tfrac{1}{2}[f(-\pi + 0) + f(\pi - 0)].$$

In particular:

A function which is smooth in the interval $[-\pi, \pi]$ can be expanded as a Fourier series in the interval $(-\pi, \pi)$. If the function has the same value at the ends of the interval, the expansion holds throughout the closed interval $[-\pi, \pi]$*.

THEOREM 2 (Dirichlet's theorem)**. **The Fourier series of a function $f(x)$ that has a finite number of extrema and is continuous in $[-\pi, \pi]$ except possibly for a finite number of points at which it has discontinuities of the first kind, is convergent at every point x of the interval to $\tfrac{1}{2}[f(x - 0) + f(x + 0)]$ (i.e. in particular, to the value of $f(x)$ itself at points where it is continuous). At both ends of the interval the sum of the series is equal to the arithmetic mean of its limits as the independent variable tends to the ends from inside the interval:**

$$\tfrac{1}{2}[(f-\pi + 0) + f(\pi - 0)].$$

In particular:

A function continuous in $[-\pi, \pi]$ that has a finite number of extrema can be expanded as a Fourier series in $(-\pi, \pi)$. If the values of the function at the ends of the interval are equal, the expansion holds throughout the closed interval $[-\pi, \pi]$.

Like Theorem 1, Dirichlet's theorem relates to a very wide class of functions. Whereas Theorem 1 allows the function to have an infinity of extremal points whilst requiring the existence of a continuous first derivative (apart from possibly a finite number of points), Theorem 2 limits the number of extrema but makes no demands regarding the existence of the derivative. Any of the functions used in analysis and its applications satisfies the conditions of Theorems 1 and 2. Since these embrace all the functions that are likely to be encountered, we can say that *every function is expressible by its Fourier series everywhere in the interval* $[-\pi, \pi]$

* If a discontinuous function $f(x)$ can be expanded as a Fourier series, we can use the $=$ sign instead of $\sim$ provided only that $f(x)$ is replaced at its discontinuities by the expression

$$\frac{f(x - 0) + f(x + 0)}{2}.$$

** G. L. DIRICHLET (1805–1859) was a famous German mathematician.

excepting at points of discontinuity and possibly the ends $x = \pm\pi$. It is curious to observe, however, that continuity is of itself insufficient to ensure that a function can be expanded as a Fourier series, and that further restrictions are necessary. Examples can be quoted of continuous functions whose Fourier series are divergent at certain points.

Trigonometric series find important applications in numerous branches of mathematics and provide very convenient methods of solution of difficult problems of mathematical physics. The theory of trigonometric series actually arose from the desire to investigate a number of problems of mechanics and physics. The later development of this important branch of analysis in fact went hand in hand with the development of mathematical physics.

The study of the expansion of functions in trigonometric series is known as harmonic analysis.

213. Fourier Series in an Arbitrary Interval. Incomplete Series.

I. THE INTERVAL $[-l, l]$. We now consider the problem of expanding as a Fourier series a function given in the interval $[-l, l]$, where l is any number. If function $f(x)$ satisfies the conditions of Theorems 1 or 2 in $[-l, l]$, the expansion is readily obtained after replacing the independent variable by $x' = \pi x/l$. This substitution leads to the expansion in $[-\pi, \pi]$ investigated above.

The Fourier series will have the form

$$f(x) = f\left(\frac{l}{\pi}x'\right) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx' + b_n \sin nx')$$

or

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos n\frac{\pi}{l}x + b_n \sin n\frac{\pi}{l}x\right),$$

where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f\left(\frac{l}{\pi}x'\right)\cos nx'\,dx' = \frac{1}{l}\int_{-l}^{l} f(x)\cos n\frac{\pi}{l}x\,dx,$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f\left(\frac{l}{\pi}x'\right)\sin nx'\,dx' = \frac{1}{l}\int_{-l}^{l} f(x)\sin n\frac{\pi}{l}x\,dx.$$

The sum of the Fourier series is here a periodic function with period $2l$.

We can therefore always confine ourselves to the interval $[-\pi, \pi]$ when studying Fourier series, since the case of any interval $[-l, l]$ reduces to this.

II. INCOMPLETE SERIES. Let $f(x)$ be expanded as a Fourier series, where $f(x)$ is an even function. The functions $f(x) \sin kx$ will now be odd, and all the coefficients b_k will vanish (see Sec. 107). *An even function thus has an incomplete Fourier series, consisting of cosines only:*

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos nx,$$

where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \cos nx \, dx = \frac{2}{\pi}\int_{0}^{\pi} f(x) \cos nx \, dx,$$

since $f(x) \cos nx$ is an even function.

Now let $f(x)$ be an odd function. The functions $f(x) \cos kx$ will now be odd, and all the coefficients a_k will vanish. Thus *an odd function has an incomplete Fourier series, consisting of sines only:*

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx,$$

where

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin nx \, dx = \frac{2}{\pi}\int_{0}^{\pi} f(x) \sin nx \, dx,$$

since $f(x) \sin nx$ is an even function.

III. ARBITRARY INTERVAL. Suppose we want to expand as a trigonometric series a function $f(x)$ in the interval $[0, \pi]$ (or what reduces to this, in any interval $[0, l]$). We can continue $f(x)$ in the interval $[-\pi, 0]$ in any manner, provided that the function $F(x)$ thus formed in the interval $[-\pi, \pi]$ (coinciding with $f(x)$ in $[0, \pi]$) satisfies the conditions of Theorem 1 or 2. Having expanded $F(x)$ as a Fourier series, we obtain the required series for $f(x)$ in the interval $[0, \pi]$ (it is of no consequence that it represents some other function, having no essential connection with $f(x)$, in the interval $[-\pi, 0]$).

In particular, $f(x)$ can be given an "*even extension*" into $[-\pi, 0]$ (i.e. the graph of $f(x)$ is "extended symmetrically" about Oy). In this case $F(x)$ is an even function and the series is incomplete (without cosines). Whereas if $f(x)$ is given an "*odd extension*" into

$[-\pi, 0]$ (i.e. the graph is extended symmetrically with respect to the origin), $F(x)$ will be an odd function, and the series will again be incomplete (with cosines only).

We reach the conclusion that, if $f(x)$ can be expanded as a trigonometric series, there exists an infinity of expansions of it in the interval $[0, \pi]$. The function can be expressed by various Fourier series, both complete and incomplete. This means that any desired number of trigonometric series can be formed, representing the same function in the interval $[0, \pi]$, and entirely different functions in $[-\pi, 0]$.

IV. SERIES IN THE COMPLEX FORM. By using complex numbers and Euler's formula, a convenient and easily remembered form can be given for the Fourier series of a function $f(x)$ in $[-\pi, \pi]$.

Let

$$f(x) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx).$$

We bring in the complex numbers $a_n - ib_n$, denoted by c_n:

$$a_n - ib_n = c_n.$$

We have, using the formulae for a_n and b_n (Sec. 211):

$$c_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) (\cos nx - i \sin nx)\, dx,$$

i.e. by Euler's formula (Sec. 135):

$$c_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, e^{-inx}\, dx. \qquad (*)$$

It may easily be observed that $a_n \cos nx + b_n \sin nx$ is the real part of the product $(a_n - ib_n)(\cos nx + i \sin nx) = c_n e^{inx}$, which can be written in turn as $\frac{1}{2}(c_n e^{inx} + \bar{c}_n e^{-inx})$, where $\bar{c}_n = a_n + ib_n$ is the conjugate of c_n. Thus

$$a_n \cos nx + b_n \sin nx = \frac{1}{2} (c_n e^{inx} + \bar{c}_n e^{-inx}).$$

Hence

$$f(x) = \frac{1}{2} a_0 + \frac{1}{2} \sum_{n=1}^{\infty} c_n e^{inx} + \frac{1}{2} \sum_{n=1}^{\infty} \bar{c}_n e^{-inx}.$$

Since $\bar{c}_n$ is got from formula (*) if we replace n by $-n$, the last equation can be written as

$$f(x) = \frac{1}{2} \sum_{n=-\infty}^{+\infty} c_n e^{inx}, \quad \text{where} \quad c_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, e^{-inx}\, dx.$$

The index of summation now runs over all integers from $-\infty$ to $+\infty$. If the factor $\frac{1}{2}$ is taken into coefficients c_n and, furthermore, we pass to any interval $[-l, l]$, the Fourier expansion can be written as

$$f(x) = \sum_{n=-\infty}^{\infty} d_n e^{in\frac{\pi}{l}x}, \quad \text{where} \quad d_n = \frac{1}{2l}\int_{-l}^{l} f(x)\, e^{-in\frac{\pi}{l}x}\, dx.$$

214. Examples. *Example* 1. Let us expand as a trigonometric series the function $f(x)$ given in the interval $[-\pi, \pi]$ by

$$f(x) = \begin{cases} ax & \text{for} \quad -\pi \leqslant x \leqslant 0, \\ bx & \text{for} \quad 0 \leqslant x \leqslant \pi, \end{cases}$$

where a and b are constants. This function obviously satisfies the conditions of Theorems 1 and 2. The graph of $f(x)$ is a step-line, made up of two pieces of the straight lines $y = ax$ and $y = bx$ (Fig. 76).

We evaluate the Fourier coefficients of $f(x)$. We have for $n \neq 0$:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\, dx$$

$$= \frac{1}{\pi}\int_{-\pi}^{0} ax\cos nx\, dx + \frac{1}{\pi}\int_{0}^{\pi} bx\cos nx\, dx$$

$$= \frac{a}{\pi}\left(\frac{x\sin nx}{n}\Bigg|_{-\pi}^{0} - \int_{-\pi}^{0}\frac{\sin nx}{n}\, dx\right) +$$

$$+ \frac{b}{\pi}\left(\frac{x\sin nx}{n}\Bigg|_{0}^{\pi} - \int_{0}^{\pi}\frac{\sin nx}{n}\, dx\right)$$

$$= \frac{a}{\pi}\,\frac{\cos nx}{n^2}\Bigg|_{-\pi}^{0} + \frac{b}{\pi}\,\frac{\cos nx}{n^2}\Bigg|_{0}^{\pi}$$

$$= \frac{a}{\pi n^2}[1 - (-1)^n] + \frac{b}{\pi n^2}[(-1)^n - 1]$$

$$= \frac{a-b}{\pi n^2}[1 - (-1)^n],$$

i.e.

$$a_1 = \frac{a-b}{\pi \cdot 1^2} \cdot 2, \quad a_2 = 0, \quad a_3 = \frac{a-b}{\pi \cdot 3^2} \cdot 2, \quad a_4 = 0, \ldots$$

We have with $n = 0$:

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{0} ax\,dx + \frac{1}{\pi}\int_{0}^{\pi} bx\,dx = \frac{b-a}{2}\,\pi.$$

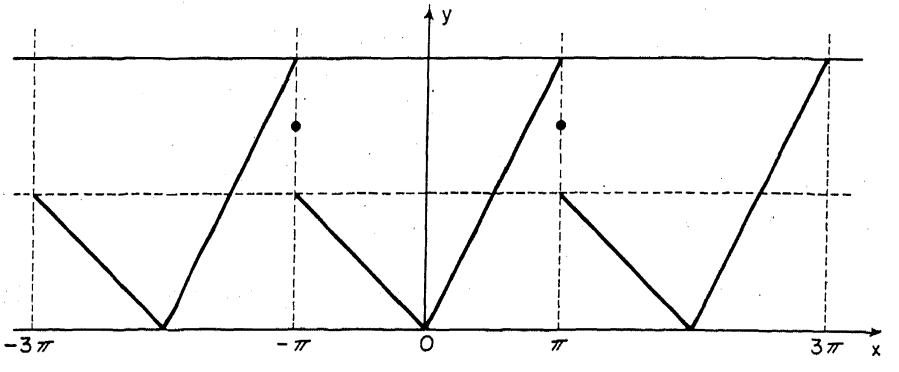

FIG. 76

Furthermore,

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx$$

$$= \frac{1}{\pi}\int_{-\pi}^{0} ax\sin nx\,dx + \frac{1}{\pi}\int_{0}^{\pi} bx\sin nx\,dx$$

$$= \frac{a}{\pi}\left(-\frac{x\cos nx}{n}\Big|_{-\pi}^{0} + \int_{-\pi}^{0}\frac{\cos nx}{n}\,dx\right) +$$

$$+ \frac{b}{\pi}\left(-\frac{x\cos nx}{n}\Big|_{-\pi}^{0} + \int_{0}^{\pi}\frac{\cos nx}{n}\,dx\right)$$

$$= \frac{a}{\pi}\,\frac{-\pi}{n}(-1)^n - \frac{b}{\pi}\,\frac{\pi}{n}(-1)^n = \frac{a+b}{n}(-1)^{n+1},$$

i.e.

$$b_1 = \frac{a+b}{1}, \quad b_2 = -\frac{a+b}{2}, \quad b_3 = \frac{a+b}{3}, \quad b_4 = -\frac{a+b}{4}, \ldots$$

The Fourier series is therefore

$$f(x) = \frac{b-a}{4}\pi + \left(\frac{a-b}{\pi}\cdot\frac{2}{1^2}\cos x + \frac{a+b}{1}\sin x\right) -$$

$$- \frac{a+b}{2}\sin 2x + \left(\frac{a-b}{\pi}\cdot\frac{2}{3^2}\cos 3x + \frac{a+b}{3}\sin 3x\right) -$$

$$- \frac{a+b}{4}\sin 4x + \cdots$$

or

$$f(x) = \frac{b-a}{4}\pi + \frac{2(a-b)}{\pi}\left(\frac{1}{1^2}\cos x + \frac{1}{3^2}\cos 3x + \cdots\right) +$$

$$+ (a+b)\left(\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \cdots\right).$$

The series represents $f(x)$ in the interval $(-\pi, \pi)$, whilst at $x = \pm\pi$ it is equal to

$$\frac{f(-\pi+0) + f(\pi-0)}{2} = \frac{-a\pi + b\pi}{2} = \pi\,\frac{b-a}{2}.$$

With $x = \pm\pi$ we get an interesting numerical series for π. We have:

$$\pi\,\frac{b-a}{2} = \frac{b-a}{4}\pi + \frac{2(b-a)}{\pi}\left(\frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots\right),$$

whence

$$\frac{\pi^2}{8} = \frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots$$

Some further interesting formulae may be obtained from the expansion for $\pi^2/8$. We put:

$$\sigma = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots,$$

$$\sigma_1 = \frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots \left(= \frac{\pi^2}{8}\right),$$

$$\sigma_2 = \frac{1}{2^2} + \frac{1}{4^2} + \frac{1}{6^2} + \cdots$$

We have:

$$\sigma = \sigma_1 + \sigma_2, \quad \sigma_2 = \frac{1}{4}\sigma = \frac{1}{4}(\sigma_1 + \sigma_2),$$

or

$$3\sigma_2 = \sigma_1, \quad \text{i.e.} \quad \sigma_2 = \frac{\pi^2}{24}.$$

Thus

$$\sigma = \frac{\pi^2}{8} + \frac{\pi^2}{24} = \frac{\pi^2}{6},$$

i.e.

$$\frac{\pi^2}{6} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots$$

Let us take some particular cases of the above expansion.

Let $a = -1$ and $b = 1$. We now have in the interval $[-\pi, \pi]$: $f(x) = |x|$ (Fig. 77). We get the expansion

$$f(x) = \frac{1}{2}\pi - \frac{4}{\pi}\left(\frac{1}{1^2}\cos x + \frac{1}{3^2}\cos 3x + \frac{1}{5^2}\cos 5x + \cdots\right),$$

which holds for any x, since $f(x)$, being continued periodically, is a continuous function throughout Ox, whilst the derivative has finite discontinuities at $x = k\pi$, $k = 0, \pm 1, \pm 2, \ldots$ The

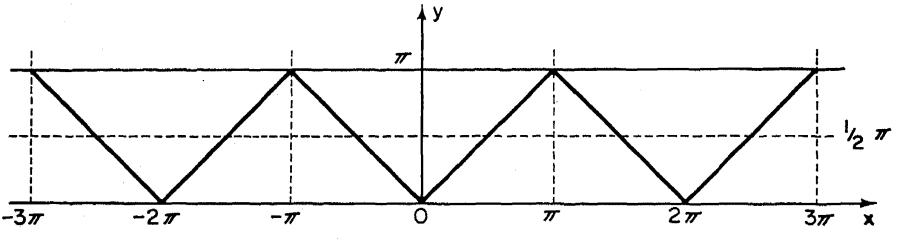

FIG. 77

function $y = |x|$ is even and can be regarded as formed by the "even extension" into the interval $[-\pi, 0]$ of $y = x$ given in the interval $[0, \pi]$.

Let $a = b = 1$. Now $f(x) = x$ (Fig. 78), and the series becomes

$$f(x) = 2\left(\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \cdots\right).$$

This equation holds for all x, $-\pi < x < \pi$, whilst its sum is zero for $x = \pm\pi$. When $x = \frac{1}{2}\pi$ we arrive at the familiar series

for the number π (obtained in Sec. 133 by expanding arc tan x as a Taylor series):

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \cdots$$

The function $f(x) = x$ is odd in $[-\pi, \pi]$ and can be regarded as formed by the "odd extension" of function $y = x$ given in the interval $[0, \pi]$.

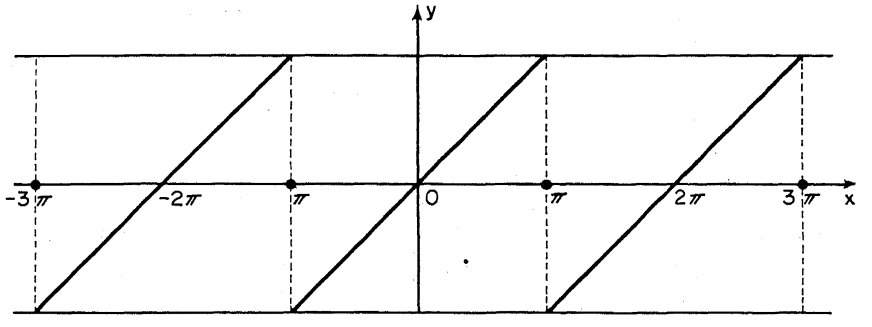

FIG. 78

It will be seen that $f(x) = x$ is represented in $[0, \pi]$ both by an "even" and by an "odd" trigonometric series:

$$x = \frac{\pi}{2} - \frac{4}{\pi}\left(\frac{1}{1^2}\cos x + \frac{1}{3^2}\cos 3x + \frac{1}{5^2}\cos 5x + \cdots\right)$$

$$= 2\left(\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \cdots\right).$$

Let $a = 1$, $b = 0$. Now,

$$f(x) = \begin{cases} x & \text{for} \quad -\pi \leqslant x \leqslant 0, \\ 0 & \text{for} \quad 0 \leqslant x \leqslant \pi \end{cases}$$

(Fig. 79). We now have the series

$$f(x) = -\frac{\pi}{4} + \frac{2}{\pi}\left(\frac{1}{1^2}\cos x + \frac{1}{3^2}\cos 3x + \frac{1}{5^2}\cos 5x + \cdots\right) +$$

$$+ \left(\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \cdots\right).$$

This is equal to x everywhere in $[-\pi, 0]$, whilst it converges to zero in $[0, \pi]$; when $x = \pm\pi$ its sum is equal to $-\frac{1}{2}\pi$.

Example 2. Let us expand as a trigonometric series the function $f(x) = 1$ given in $[0, \pi]$. We can extend it into $[-\pi, 0]$ in any manner, and in particular, in an "odd" manner (Fig. 80). As is easily seen, the Fourier series reduces to its zero term and we get the identity $1 = 1$ when the function has an "even" extension.

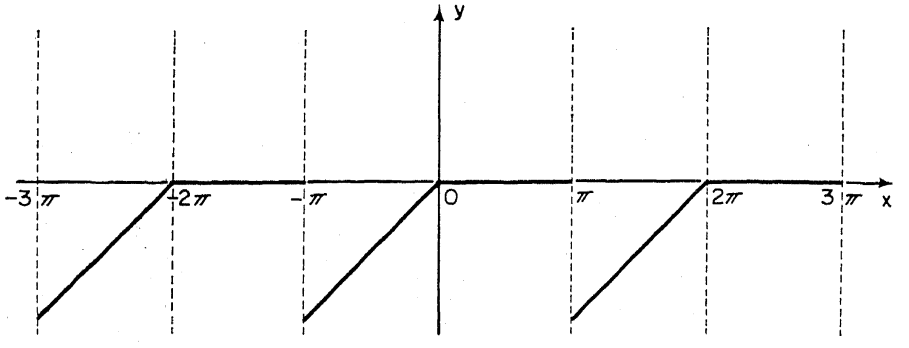

FIG. 79

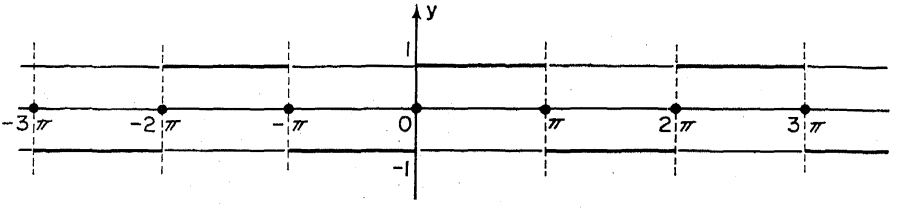

FIG. 80

Let us take the odd extension. Now,

$$b_n = \frac{2}{\pi}\int_0^{\pi} \sin nx\, dx = \frac{2}{\pi} \cdot \frac{-\cos nx}{n}\Big|_0^{\pi} = \frac{2}{\pi n}\,[1 - (-1)^n],$$

i.e. $$b_1 = \frac{4}{\pi \cdot 1}, \quad b_2 = 0, \quad b_3 = \frac{4}{\pi \cdot 3}, \ldots$$

We therefore get, for $0 < x < \pi$:

$$1 = \frac{4}{\pi}\left(\sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \cdots\right),$$

whence

$$\frac{\pi}{4} = \sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \cdots, \quad 0 < x < \pi.$$

In $(-\pi, 0)$ the sum of the series on the right is everywhere equal to $-\pi/4$, whilst it vanishes at the points $x = -\pi$, $x = 0$, $x = \pi$.

In particular, when $x = \frac{1}{2}\pi$ we again get the familiar series

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \cdots$$

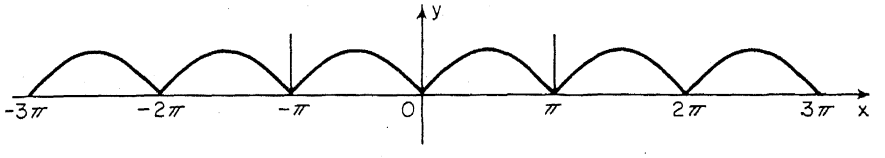

FIG. 81

Example 3. Let us find the Fourier series of $f(x) = |\sin x|$ (Fig. 81). This function is obtained by the "even extension" of $\sin x$ from $[0, \pi]$ to $[-\pi, 0]$. Hence

$$a_0 = \frac{\pi}{2}\int_0^{\pi} \sin x\, dx = \frac{4}{\pi},$$

$$a_n = \frac{2}{\pi}\int_0^{\pi} \sin x \cos nx\, dx = \frac{1}{\pi}\int_0^{\pi} [\sin (n+1)x - \sin (n-1)x]\, dx$$

$$= \begin{cases} 0, & \text{if } n \text{ is odd} \\ \dfrac{-4}{\pi(n^2-1)}, & \text{if } n \text{ is even.} \end{cases}$$

Consequently,

$$|\sin x| = \frac{2}{\pi} - \frac{4}{\pi}\left(\frac{1}{1\,.\,3}\cos 2x + \frac{1}{3\,.\,5}\cos 4x + \frac{1}{5\,.\,7}\cos 6x + \cdots \right.$$

$$\left. \cdots + \frac{1}{(2n-1)(2n+1)}\cos 2nx + \cdots\right).$$

The expansion holds for any x. In particular, when $x = 0$ we get the relationship

$$\frac{1}{2} = \frac{1}{1\,.\,3} + \frac{1}{3\,.\,5} + \frac{1}{5\,.\,7} + \cdots + \frac{1}{(2n-1)(2n+1)} + \cdots,$$

which can also be proved by elementary methods.

215. Uniform Convergence of Fourier Series. Convergence "in mean."

I. UNIFORM CONVERGENCE. The sufficient conditions for convergence of a Fourier series to a given function were given in Theorems 1 and 2 (Sec. 212). But we also want to know the conditions in which the convergence is uniform, since it is only then that we can guarantee that the same Fourier polynomial is an approximate expression for a function with the same degree of error throughout the interval. It is clear that the Fourier series cannot be uniformly convergent in the neighbourhood of a discontinuity, since otherwise its sum would be a continuous function in accordance with the general theorem of Sec. 127.

THEOREM. **The Fourier series of a function $f(x)$ which is continuous and has a piecewise smooth derivative in $[-\pi, \pi]$ and has equal values $f(-\pi) = f(\pi)$ at the ends is uniformly convergent to $f(x)$ throughout $[-\pi, \pi]$.**

Proof. We write down the expressions for the Fourier coefficients:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx, \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx$$

and integrate by parts:

$$a_n = \frac{1}{\pi} f(x)\frac{\sin nx}{n}\Bigg|_{-\pi}^{\pi} - \frac{1}{\pi n}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx$$

$$= -\frac{1}{\pi n}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx = -\frac{b_n'}{n},$$

$$b_n = -\frac{1}{\pi} f(x)\frac{\cos nx}{n}\Bigg|_{-\pi}^{\pi} + \frac{1}{\pi n}\int_{-\pi}^{\pi} f'(x)\cos nx\,dx$$

$$= \frac{1}{\pi n}\int_{-\pi}^{\pi} f'(x)\cos nx\,dx = \frac{a_n'}{n},$$

since $f(-\pi) = f(\pi)$. The a_n' and b_n' are the Fourier coefficients of $f'(x)$. In view of the condition that $f'(x)$ is piecewise smooth, we can apply the theorem of Sec. 211, II, so that the series

$$\sum_{n=1}^{\infty}\frac{|a_n|}{n} = \sum_{n=1}^{\infty}|b_n| \quad \text{and} \quad \sum_{n=1}^{\infty}\frac{|b_n'|}{n} = \sum_{n=1}^{\infty}|a_n|$$

are convergent; and since $|a_n \cos nx + b_n \sin nx| \leqslant |a_n| + |b_n|$, the Fourier series

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)$$

is uniformly convergent by Weierstrass's test (Sec. 127) (and is absolutely convergent), its sum being the given $f(x)$, as follows from Theorem 1 of Sec. 212.

For instance, the series

$$\frac{1}{2}\pi - \frac{4}{\pi}\left(\frac{1}{1^2}\cos x + \frac{1}{3^2}\cos 3x + \frac{1}{5^2}\cos 5x + \cdots\right)$$

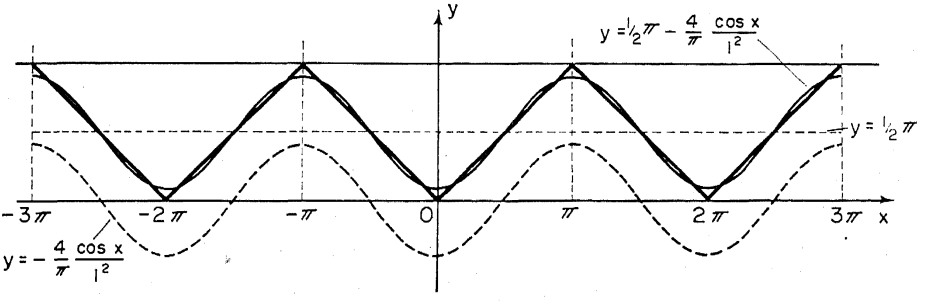

FIG. 82

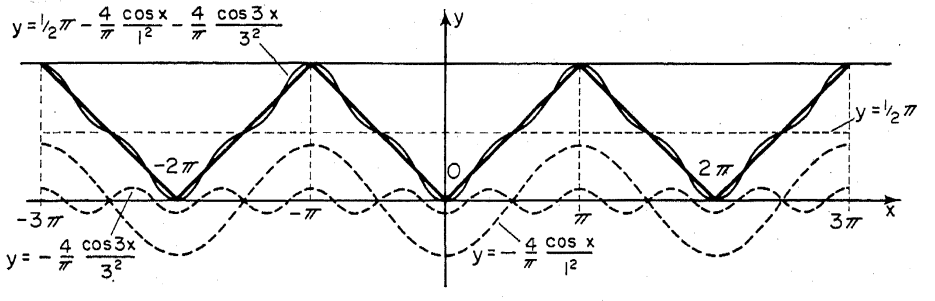

FIG. 83

(Sec. 214, Example 1) is uniformly convergent* to the "step" function ($y = |x|$) whose graph is shown in Fig. 77. Figures 82 and 83 illustrate the uniformity of the approximation of the Fourier polynomials to the required function; successive harmonics of

* The uniform convergence of this series is also easily proved directly. In fact, its coefficients form the convergent series

$$\frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots$$

the series are shown dotted, whilst the full line is their sum, i.e. the graph of the corresponding Fourier polynomials.

The following general theorem can be proved with the aid of some additional arguments.

THEOREM. **The Fourier series of a piecewise smooth function having a piecewise smooth derivative, is uniformly convergent to the function throughout any interval not containing a discontinuity of the function*.**

In addition to the fact that Fourier polynomials can, given suitable conditions, approximate a given function as accurately and

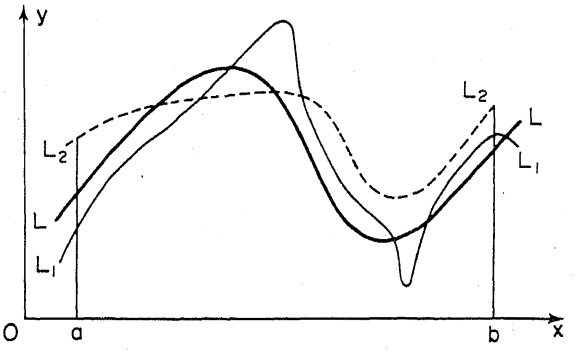

FIG. 84

uniformly as required, they have an important further property indicating their "closeness" to the approximated function.

A new concept must be introduced in order to explain this property; this is the concept of mean square deviation, which has important applications in various problems of analysis.

II. CONVERGENCE "IN MEAN". We take two functions $F(x)$ and $F_1(x)$, integrable along with their squares in the interval $[a, b]$. We impose no further restrictions whatever on these functions. If we regard $F_1(x)$ as an "approximation" to $F(x)$ in $[a, b]$, i.e. replace $F(x)$ by $F_1(x)$, the resulting error or "deviation" is

$$R = F(x) - F_1(x),$$

which can be characterized with the aid of various quantities (tests). Up till now we have characterized this deviation either by its order of smallness in the neighbourhood of a point (as was done for Taylor polynomials) or by the maximum of its modulus in the

* There are even subtler theorems on the uniform convergence of Fourier series, though we must omit them (see G.M. FIKHTENGOL'TS, *Course of Differential and Integral Calculus, (Kurs differentsial'nogo i integral'nogo ischisleniya)*, vol. III, p. 586, Gost. 1949).

interval (as in the case of Chebyshev polynomials). A further important quantity that may be used to indicate the deviation is its root mean square, i.e.

$$\Delta = \sqrt{\frac{1}{b-a}\int_a^b R^2\,dx}\,.$$

Let us examine the meaning of this quantity. Squaring R eliminates the difference between positive and negative deviations of $F_1(x)$ from $F(x)$*, which in fact usually have no effect on the closeness of an approximation; the arithmetic mean of R^2 yields a general characteristic deviation which is often more natural than other characteristics. It is sometimes more important to know how little one function deviates from another "on the whole"—in the sense of a mean value—than to know say the greatest of all possible local (point) deviations. Fig. 84 illustrates the graphs L_1 (thin full line) and L_2 (thin dotted line) of two functions approximating to the function whose graph is L (thick full line). If we estimate the approximation in terms of the greatest deviation, the second function (graph L_2) represents the given function better than the first (graph L_1) inasmuch as there are individual pieces of the interval $[a, b]$ in which the first function deviates strongly (more so than the second) from the given function. This conclusion could not be altered even by the complete merging of the first with the given function at all remaining parts of the interval. Nevertheless, from the point of view of the closeness of the functions throughout the interval, which may well be expressed by Δ, there is no doubt that, conversely, the first function represents the given function better than does the second.

Definition. **We call Δ the *root mean square deviation* of function $F_1(x)$ from function $F(x)$ in the interval $[a, b]$.**

If we regard the approximation of function $F_1(x)$ to the given $F(x)$ in $[a, b]$ as being the better, the smaller Δ, we arrive at the statement of the problem of approximations in which we seek among the class of approximating functions the function giving the least root mean square deviation Δ. Such an approximation is described as "*best in the mean square sense*", whilst the method of distinguishing this "best" function is often termed the "*method of least squares*".

* This can be achieved by using $|R|$ as well as R^2, but the modulus does not lead to as convenient and simple a relationship as that obtained by taking the square.

THEOREM. **Among the trigonometric polynomials of order n, the best approximation to a given function in the mean square sense is the n-th order Fourier polynomial of this function,**
i.e. in other words,

the mean square deviation of a trigonometric polynomial of order n from a given function has its least value for the Fourier polynomial of the function.

Proof. Let the function be integrable together with its square in the interval concerned, which we shall always take to be $[-\pi, \pi]$.

Let $F_n(x)$ be any n-th order polynomial:

$$F_n(x) = \frac{1}{2}\alpha_0 + \sum_{k=1}^{n} (\alpha_k \cos kx + \beta_k \sin kx)$$

and let its root mean square deviation from the given function $f(x)$ be Δ'_n (it is more convenient to take Δ'^2_n than Δ'_n):

$$\Delta'^2_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} [f(x) - F_n(x)]^2 \, dx .$$

Here, α_k, β_k are arbitrary numbers. We now show that Δ'_n (or what amounts to the same thing, Δ'^2_n) attains its minimum value when $\alpha_k = a_k$, $\beta_k = b_k$ $(k = 0, 1, \ldots, n)$, where a_k, b_k are the Fourier coefficients of $f(x)$.

On repeating the working of Sec. 211, we in fact get

$$\Delta'^2_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f^2(x)\, dx - \left[\frac{1}{2}\alpha_0 a_0 + \sum_{k=1}^{n} (\alpha_k a_k + \beta_k b_k)\right] + \\ + \frac{1}{2}\left[\frac{1}{2}\alpha_0^2 + \sum_{k=1}^{n} (\alpha_k^2 + \beta_k^2)\right],$$

whence

$$2\,\Delta'^2_n = \left\{\frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\, dx - \left[\frac{1}{2}a_0^2 + \sum_{k=1}^{n} (a_k^2 + b_k^2)\right]\right\} + \\ + \frac{1}{2}(\alpha_0 - a_0)^2 + \sum_{k=1}^{n} [(\alpha_k - a_k)^2 + (\beta_k - b_k)^2] .$$

The expression in the curly brackets is the same for any choice of polynomial $F_n(x)$. We put:

$$M = \frac{1}{2}(\alpha_0 - a_0)^2 + \sum_{k=1}^{n} [(\alpha_k - a_k)^2 + (\beta_k - b_k)^2] .$$

Since $M \geqslant 0$, Δ_n' attains its least value with $M = 0$, which is only possible if $\alpha_k = a_k$ and $\beta_k = b_k$, $k = 0, 1, \ldots, n$.

This is what we wished to prove.

There is thus seen to be a different possible approach to trigonometric series to the one that we have been using. Instead of starting from the problem of finding the n-th order trigonometric polynomial whose deviation tends to zero ($R_n \to 0$), we could start from the problem of finding the polynomials giving the best approximation in "mean." We should in fact again arrive at a Fourier series.

216. The Parseval-Lyaponov Theorem.

Definition. A sequence of functions $F_n(x)$ is said to tend to function $F(x)$ in the mean square sense (or simply "in mean") if the mean square deviation of the functions from $F(x)$ in the interval $[a, b]$ tends to zero as $n \to \infty$:

$$\lim_{n\to\infty} \frac{1}{b-a} \int_a^b [F(x) - F_n(x)]^2 \, dx = 0 .$$

It must be borne in mind that convergence "in mean" does not necessarily involve ordinary (point) convergence to the same function*.

As was shown in Sec. 215, I, *the Fourier series for function $f(x)$, continuous in $(-\pi, \pi)$ with piecewise smooth derivative is convergent* (in the ordinary "point" sense) *to its generating function, the convergence being uniform in any closed interval contained in $(-\pi, \pi)$.* The following theorem also holds for such a function.

THEOREM. **The Fourier series of a function $f(x)$, continuous in $(-\pi, \pi)$ and with a piecewise smooth derivative, is convergent "in mean" to $f(x)$, i.e. the corresponding Fourier polynomials $\Phi_n(x)$, of**

* Let us take say the sequence of functions

$$F_n(x) = \sqrt{\frac{1}{1 + (nx)^2}} .$$

This sequence is convergent "in mean" to zero in $[-1, 1]$, since

$$\lim_{n\to\infty} \frac{1}{2} \int_{-1}^{1} \frac{1}{1 + (nx)^2} \, dx = 0 .$$

Yet the sequence does not tend to zero in the ordinary sense, since $F_n(0) = 1$.

order n, tend to $f(x)$ in the mean square sense as $n \to \infty$; the formula holds:

$$\tfrac{1}{2}a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2) = \frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\,dx, \qquad (*)$$

this being known as *Parseval's formula.*

Proof. We have (see Sec. 211, II):

$$R_n = f(x) - \Phi_n(x)$$

and

$$\Delta_n^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} R_n^2\,dx = \frac{1}{2}\left\{\frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\,dx - \left[\frac{1}{2}a_0^2 + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right]\right\}.$$

As n increases new negative terms are added in the expression for Δ_n^2, so that Δ_n diminishes. Moreover, it is easily shown that $\Delta_n \to 0$ as $n \to \infty$, i.e. given ε, an $n = N$ can be chosen such that $|R_n| < \varepsilon$ for $n > N$ and any x. Consequently,

$$\Delta_n^2 < \frac{1}{2\pi} \cdot 2\pi\varepsilon^2, \quad \text{or} \quad |\Delta_n| < \varepsilon,$$

i.e. $\lim\limits_{n\to\infty} \Delta_n = 0$, which implies formula (*). This is what we had to prove.

There is an interesting corollary to Parseval's theorem.

THEOREM. **Two different continuous functions cannot have the same Fourier series, i.e. a continuous function is uniquely determined by its Fourier coefficients.**

Proof. Let the two functions $f_1(x)$ and $f_2(x)$ have the same Fourier coefficients. The Fourier coefficients of $f(x) = f_1(x) - f_2(x)$ are now obviously zero, so that, by Parseval's theorem,

$$\int_{-\pi}^{\pi} f^2(x)\,dx = 0,$$

whence it follows (see Sec. 90) that $f(x) \equiv 0$, i.e. $f_1(x) = f_2(x)$. This is what we had to show.

Consequently, *if a Fourier series converges, it must converge to its generating* (continuous) *function.*

Parseval's formula holds in more general conditions.

PARSEVAL–LYAPUNOV THEOREM. **The Fourier series of any $f(x)$ which is integrable along with its square in $(-\pi, \pi)$ converges "in**

mean" to $f(x)$, i.e. $\lim_{n\to\infty} \Delta_n = 0$, so that Parseval's theorem holds for such a function.

A strict proof of this theorem was first given by A. M. LYAPUNOV (see Introduction, Sec. 4).

Each of the theorems discussed in this chapter is stated for definite conditions imposed on the properties of the functions with which the theorem is concerned. Mathematicians have made great efforts to discover the minimum conditions needed, and a substantial specialist literature has grown up around the various problems of the theory of trigonometric series.

A considerable number of Soviet scientists have played a decisive role in developing this theory and its practical applications. It is impossible for us here to describe the theorems in the subtle and perfected form which has now been attained.

We shall summarize the general theory of trigonometric series by stating the basic results in the simplest conditions, which are readily memorized and are most often encountered in practice.

If a function $f(x)$ is continuous in the interval $[-\pi, \pi]$ and has a piecewise smooth derivative, whilst also $f(-\pi) = f(\pi)$ (i.e. $f(x)$ remains continuous on being extended periodically throughout Ox), **its Fourier series is uniformly convergent to $f(x)$ in $[-\pi, \pi]$; it is also convergent "in mean" to $f(x)$. Thus**

$$\lim_{n\to\infty} \Phi_n(x) = f(x)$$

for all x, $-\pi \leqslant x \leqslant \pi$, whilst for any previously assigned $\varepsilon > 0$ there is a corresponding N such that, for all $n > N$ and any x of $[-\pi, \pi]$ we have

$$|f(x) - \Phi_n(x)| < \varepsilon.$$

Moreover,

$$\lim_{n\to\infty} \Delta_n^2 = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} [f(x) - \Phi_n(x)]^2 \, dx = 0,$$

which is equivalent to Parseval's formula

$$\frac{1}{2} a_0^2 + \sum_{n=1}^{\infty} (a_n^2 + b_n^2) = \frac{1}{\pi} \int_{-\pi}^{\pi} f^2(x) \, dx.$$

($\Phi_n(x)$ is the n-th order Fourier polynomial of $f(x)$, a_n, b_n, $n = 0, 1, 2, \ldots$ are its Fourier coefficients.)

If $f(x)$ has a finite number of discontinuities of the first kind in a finite interval, the only modifications are: at a discontinuity the series is convergent to $\frac{1}{2}[f(x-0)+f(x+0)]$, whilst the convergence is uniform in any interval not containing a discontinuity.

3. Krylov's Method. Practical Harmonic Analysis

217. Order of the Coefficients. As we know (Sec. 211, II), the coefficients of the Fourier series tend to zero. It is an advantage from the point of view of evaluations if they tend to zero as "fast" as possible, so that the remainder term becomes as small as possible, i.e. the convergence is "improved." If the coefficients are too big the series converges slowly, and we have to try to improve the convergence by carrying out transformations to make the coefficients smaller.

THEOREM. *Given a continuous function $f(x)$ in $[-\pi, \pi]$ such that $f(-\pi) = f(\pi)$ and the derivatives $f'(x), \ldots, f^{(p-1)}(x)$ exist everywhere, whilst*

$$f'(-\pi) = f'(\pi), \ldots, f^{(p-1)}(-\pi) = f^{(p-1)}(\pi),$$

and $f^{(p)}(x)$ is integrable together with its square, the Fourier coefficients a_n and b_n will be infinitesimals of higher order than the p-th with respect to $1/n$ as $n \to \infty$, i.e.

$$\frac{a_n}{\frac{1}{n^p}} = n^p a_n \to 0, \quad \frac{b_n}{\frac{1}{n^p}} = n^p b_n \to 0.$$

Proof. We consider Fourier coefficients a_n and b_n:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx, \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx;$$

we integrate by parts:

$$a_n = \frac{1}{\pi}\left(\frac{f(x)\sin nx}{n}\right)\Bigg|_{-\pi}^{\pi} - \frac{1}{\pi n}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx = -\frac{b_n'}{n},$$

$$b_n = \frac{1}{\pi}\left(-\frac{f(x)\cos nx}{n}\right)\Bigg|_{-\pi}^{\pi} + \frac{1}{\pi n}\int_{-\pi}^{\pi} f'(x)\cos nx\,dx = \frac{a_n'}{n},$$

where a_n', b_n' are the n-th order Fourier coefficients of $f'(x)$. By the same relationships, $b_n' = a_n''/n$, $a_n' = -b_n''/n$, so that

$$a_n = -\frac{a_n''}{n^2}, \quad b_n = -\frac{b_n''}{n^2},$$

where a_n'', b_n'' are the n-th order Fourier coefficients of $f''(x)$. By proceeding in this way, we arrive at the formulae

$$a_n = \pm\frac{a_n^{(p)}}{n^p}\left(\text{or } \pm\frac{b_n^{(p)}}{n^p}\right), \quad b_n = \pm\frac{b_n^{(p)}}{n^p}\left(\text{or } \pm\frac{a_n^{(p)}}{n^p}\right), \qquad (*)$$

where $a_n^{(p)}$, $b_n^{(p)}$ are the n-th order Fourier coefficients of $f^{(p)}(x)$. Since, by the theorem of Sec. 211, $a_n^{(p)} \to 0$, $b_n^{(p)} \to 0$ as $n \to \infty$, our theorem follows from (*).

Now let $f(x)$ have a discontinuity of the first kind in $[-\pi, \pi]$ at x_1, $-\pi \leqslant x_1 \leqslant \pi$. Then

$$a_n = \frac{1}{\pi}\int_{-\pi}^{x_1} f(x)\cos nx\,dx + \frac{1}{\pi}\int_{x_1}^{\pi} f(x)\cos nx\,dx$$

$$= \frac{1}{\pi}\frac{f(x)\sin nx}{n}\Big|_{-\pi}^{x_1} - \frac{1}{\pi n}\int_{-\pi}^{x_1} f'(x)\sin nx\,dx + \frac{1}{\pi}\frac{f(x)\sin nx}{n}\Big|_{x_1}^{\pi} -$$

$$- \frac{1}{\pi n}\int_{x_1}^{\pi} f'(x)\sin nx\,dx$$

$$= \frac{1}{\pi}\frac{f(x_1 - 0) - f(x_1 + 0)}{n}\sin nx_1 - \frac{1}{\pi n}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx$$

$$= -\frac{1}{\pi}\frac{\delta(x_1)}{n}\sin nx_1 - \frac{b_n'}{n},$$

where $\delta(x_1)$ is the jump of $f(x)$ at x_1, i.e.

$$\delta(x_1) = f(x_1 + 0) - f(x_1 - 0).$$

Similarly,

$$b_n = \frac{1}{\pi}\frac{\delta(x_1)}{n}\cos nx_1 + \frac{a_n'}{n}.$$

If $f(x)$ has discontinuities of the first kind at $x_1, x_2, \ldots, x_m$ in $[-\pi, \pi]$, the Fourier coefficients a_n and b_n take the form

$$\left.\begin{aligned} a_n &= -\frac{1}{\pi}\frac{\sum_{k=1}^{m}\delta(x_k)\sin nx_k}{n} - \frac{b_n'}{n}, \\ b_n &= \frac{1}{\pi}\frac{\sum_{k=1}^{m}\delta(x_k)\cos nx_k}{n} + \frac{a_n'}{n}, \end{aligned}\right\} \qquad (**)$$

where $\delta(x_k) = f(x_k + 0) - f(x_k - 0)$ is the jump of the function at x_k.

The first terms in the expressions for a_n and b_n are infinitesimals (as $n \to \infty$) of the same order as $1/n$, whilst the second terms are of higher order. Thus a_n and b_n are of the same order as $1/n$. This involves very slow convergence of the series. If the function has even one discontinuity inside the interval or at one end (i.e. the limits $f(-\pi + 0)$ and $f(\pi - 0)$ are not the same) the size of its Fourier coefficients is considerably affected; even though the function is smooth everywhere apart from the one discontinuity, the

presence of the latter still leads to a lowering of the order of the coefficients. If the coefficients are of the same order as $1/n$, the series is not suitable for practical evaluations, since to achieve an accuracy of 0.01 in the approximation would require several tens of terms of the series. In addition, the need quite often arises in practical problems for differentiation of the Fourier series, and each differentiation lowers the order of the coefficients by unity. For, if

$$f(x) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

then

$$f'(x) = \sum_{n=1}^{\infty} (-n a_n \sin nx + n b_n \cos nx),$$

so that, if a_n and b_n are of the p-th order of smallness, i.e. $n^p a_n \to A \neq 0$, $n^p b_n \to B \neq 0$, the coefficients of the "derived" series $n a_n$ and $n b_n$ will only be of the $(p-1)$-th order. In particular, when the function has a discontinuity (when the coefficients are of the first order of smallness) it is impossible to differentiate the Fourier series term by term, since this operation leads to a divergent series (the coefficients do not tend to zero)*.

We can now give precision to the theorem mentioned on p. 350. Let discontinuities of the first kind be encountered only in the p-th order derivative; we now see, as above, that coefficients a_n and b_n are infinitesimals of order $(p+1)$ with respect to $1/n$. For, by the theorem just proved, coefficients $a_n^{(p)}$, $b_n^{(p)}$ of the function $f^{(p)}(x)$ are of the first order, so that, by expression (*), a_n and b_n are of order $(p+1)$ with respect to $1/n$.

Their general form is

$$a_n = -\frac{1}{\pi} \frac{\sum_{k=1}^{m} \delta^{(p)}(x_k) \sin nx_k}{n^{p+1}} - \frac{b_n^{(p+1)}}{n^{p+1}},$$

$$b_n = \frac{1}{\pi} \frac{\sum_{k=1}^{m} \delta^{(p)}(x_k) \cos nx_k}{n^{p+1}} + \frac{a_n^{(p+1)}}{n^{p+1}},$$

where $x_1, x_2, \ldots, x_m$ are the points at which $f^{(p)}(x)$ is discontinuous; $\delta^{(p)}(x_k)$ is the jump of this function at x_k; $a_n^{(p+1)}$, $b_n^{(p+1)}$ are the n-th order Fourier coefficients of $f^{(p+1)}(x)$.

The converse also holds: if a_n and b_n are of order $(p+1)$ with respect to $1/n$, the p-th order derivative $f^{(p)}(x)$ has a discontinuity in $[-\pi, \pi]$.

* It is worth remarking that the possibility of term-by-term differentiation of a Fourier series is decided from the general rule of Sec. 128. As regards integration, this is in fact always permissible term by term for a convergent Fourier series; the series thus obtained is convergent to the corresponding integral of the function represented by the series. We shall leave the proof aside, but draw the reader's attention to the fact which makes our assertion almost obvious: integration raises by unity the order of smallness of the coefficients. In general, every differentiation of a Fourier series worsens its convergence, and every integration improves it.

We thus arrive at the proposition:

The necessary and sufficient condition for the derivative $f^{(p)}(x)$ *to have a discontinuity in* $[-\pi, \pi]$ *is that the Fourier coefficients* a_n *and* b_n *of* $f(x)$ *be of order* $(p+1)$ *with respect to* $1/n$ *as* $n \to \infty$.

It follows from this, by the way, that the necessary and sufficient conditions for $f(x)$ to be infinitely differentiable in $[-\pi, \pi]$ are that

$$\lim_{n\to\infty} n^p a_n = 0, \quad \lim_{n\to\infty} n^p b_n = 0$$

for any p i.e. **the order of smallness of the Fourier coefficients of $f(x)$ with respect to $1/n$ must be "infinite".**

218. Krylov's Method of Improving the Convergence of Trigonometric Series.

It is only satisfactory to carry out computations with trigonometric series if the order of the coefficients is not lower than two. If this is not the case or if the coefficients do not in general satisfy this condition, it is necessary to improve the convergence, by using a method proposed for this purpose by A. N. KRYLOV (see Introduction, Sec. 5) in his celebrated work *Some Differential Equations of Mathematical Physics with Engineering Applications (O nekotorykh differentsial'nykh uravneniyakh matematicheskoi fiziki, imeyushchikh prilozheniya v tekhnicheskikh voprosakh)*. The essence of the method lies in subtracting from the given function as simple a function as possible with the same peculiarities. This eliminates the causes of the poor convergence, the difference being a function whose Fourier coefficients are of higher orders of smallness.

We shall dwell a little on this problem whilst referring the reader for further details to Krylov's original article.

Suppose we have the convergent trigonometric series

$$f(x) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

where coefficients a_n and b_n are of the first order of smallness with respect to $1/n$; we separate the "principal parts", proportional to $1/n$, from a_n and b_n:

$$a_n = \frac{A_n}{n} + r_n, \quad b_n = -\frac{B_n}{n} + \varrho_n,$$

where r_n and ϱ_n are of higher orders of smallness than $1/n$.

These principal parts define the discontinuities of $f(x)$ in $[-\pi, \pi]$; it is sometimes possible to find the relevant points due to our knowledge of the form of coefficients A_n and B_n (formula (**), Sec. 217):

$$A_n = -\frac{1}{\pi} \sum_{k=1}^{m} \delta(x_k) \sin nx_k, \quad B_n = \frac{1}{\pi} \sum_{k=1}^{m} \delta(x_k) \cos nx_k. \qquad (*)$$

Suppose we have found x_k and $\delta(x_k)$. We now take the "piecewise linear" function $\varphi_1(x)$ having the same jumps $\delta(x_k)$ at the points x_k and expand it as a Fourier series. The graph of $\varphi_1(x)$ consists of a number of segments of

straight lines. This function so to speak "takes up" all the discontinuities of $f(x)$, and the difference

$$f_1(x) = f(x) - \varphi_1(x)$$

is a continuous function, whose Fourier coefficients are infinitesimals of order not less than two with respect to $1/n$.

Finding $\varphi_1(x)$ usually implies actual summation of the series

$$\sum_{n=1}^{\infty} \left(\frac{A_n}{n} \cos nx + \frac{B_n}{n} \sin nx \right),$$

i.e. expressing its sum in a finite form by means of elementary functions.

On subtracting the Fourier series of $\varphi_1(x)$ from the given trigonometric series we get

$$f_1(x) = f(x) - \varphi_1(x) = \sum_{n=1}^{\infty} (a_n^{(1)} \cos nx + b_n^{(1)} \sin nx),$$

where

$$a_n^{(1)} = \frac{A_n^{(1)}}{n^2} + r_n^{(1)}, \quad b_n^{(1)} = \frac{B_n^{(1)}}{n^2} + \varrho_n^{(1)},$$

$r_n^{(1)}$ and $\varrho_n^{(1)}$ being of higher orders of smallness than $1/n^2$.

The required function $f(x)$ can be written as the sum of two terms:

$$f(x) = \Phi_1(x) + F_1(x),$$

the first of which, i.e. $\Phi_1(x) = \varphi_1(x)$, is the known piecewise linear function, and the second, i.e. $F_1(x) = f_1(x)$ is the trigonometric series with coefficients of orders not lower than two with respect to $1/n$. We have thus reduced a trigonometric series with slow convergence (first order coefficients) to another trigonometric series with improved convergence (second order coefficients).

If second order coefficients are not satisfactory we must continue the process of improving the convergence. This is done by differentiating the series for $f_1(x)$:

$$f_1'(x) = \sum_{n=1}^{\infty} (-n a_n^{(1)} \sin nx + n b_n^{(1)} \cos nx).$$

If $A_n^{(1)}$ and $B_n^{(1)}$ do not tend to zero as $n \to \infty$, this series has the same character as the given series: its coefficients are of the first order with respect to $1/n$. We now apply to $f_1'(x)$ the same process as applied above to $f(x)$, i.e. we distinguish a piecewise linear function $\varphi_2(x)$ which "takes up" all the singularities of $f_1'(x)$, which, as is readily seen, are the singularities of $f'(x)$. We get the function

$$f_2(x) = f_1'(x) - \varphi_2(x),$$

whose Fourier coefficients are of not lower than the second order with respect to $1/n$, i.e. function $\int f_2(x)\,dx$ has coefficients of not lower than the third order. All in all, $f(x)$ is now written as the sum of three terms:

$$f(x) = \Phi_1(x) + \int \varphi_2(x)\,dx + \int f_2(x)\,dx = \Phi_1(x) + \Phi_2(x) + F_2(x),$$

the first of which, i.e. $\Phi_1(x)$ is a known piecewise linear function, the second, i.e. $\Phi_2(x) = \int \varphi_2(x)\,dx$ is a known piecewise square* function, whilst the third, i.e. $F_2(x) = \int f_2(x)\,dx$ is a trigonometric series with coefficients of order not less than three with respect to $1/n$.

We have thus further improved the convergence.

By continuing the process we can gradually reduce a given trigonometric series to a series whose coefficients have a previously assigned order of smallness (with respect to $1/n$). We have after k steps:

$$f(x) = \Phi_1(x) + \Phi_2(x) + \cdots + \Phi_k(x) + F_k(x),$$

where $\Phi_1(x)$ is a piecewise linear function, $\Phi_2(x)$ a continuous piecewise square function, ..., $\Phi_k(x)$ a continuous piecewise-polynomial (of degree k)** function, $F_k(x)$ a trigonometric series whose coefficients are of not lower than the $(k+1)$-th order with respect to $1/n$.

We can tackle similarly the problem of improving the convergence of a series, when it has coefficients of order p to start with and we wish to increase this order further.

We can say that *Krylov's method of improving the convergence of a trigonometric series consists in knocking off a part with a known sum from the given slowly convergent series in such a way that the remaining part converges fairly rapidly.*

We are mostly concerned in practice with incomplete series in the interval $[0, \pi]$ (or in general in $[0, l]$). The process of improving their convergence remains the same as described above.

219. Examples. *Example 1.* Let us improve the convergence of

$$f(x) = \sin x + \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x + \cdots + \frac{1}{n}\sin nx + \cdots$$

Here the coefficients $b_n = 1/n$ are of the first order, so that the $f(x)$ represented by this series has a discontinuity. We have by expression (*) of Sec. 218:

$$\frac{1}{\pi}\sum_{k=1}^{m}\delta(x_k)\cos nx_k = 1.$$

These relationships will be satisfied for any n if we put $m = 1$, $x_1 = 0$, $\delta(0) = \pi$; in this case the auxiliary piecewise linear function $\Phi_1(x)$ has one discontinuity in $[-\pi, \pi]$ at $x = 0$ with a jump equal to π. Since

$$\frac{f(0+0) + f(0-0)}{2} = 0, \quad \text{we have} \quad f(0+0) = \frac{\pi}{2}, \quad f(0-0) = -\frac{\pi}{2};$$

* The result of integrating a piecewise linear function is a continuous piecewise square function, i.e. such that, if $[-\pi, \pi]$ is divided into the corresponding sub-intervals, it is quadratic in each of them.

** i.e. a function such that, if $[-\pi, \pi]$ is divided into the corresponding sub-intervals, the function is a polynomial of degree k in each of them.

on further observing that $f(\pi) = 0$, we find $\Phi_1(x)$, $\Phi_1(-x) = -\Phi_1(x)$ as (Fig. 85):

$$\Phi_1(x) = \begin{cases} -\dfrac{\pi + x}{2}, & -\pi < x < 0, \\ \dfrac{\pi - x}{2}, & 0 < x < \pi. \end{cases}$$

We expand $\Phi_1(x)$ as a Fourier series. Bearing in mind that $\Phi_1(x)$ is an odd function, we find that

$$b_n = \frac{2}{\pi}\int_0^{\pi} \frac{\pi - x}{2} \sin nx\, dx = \frac{1}{n}$$

i.e.

$$\Phi_1(x) = \sin x + \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x + \cdots + \frac{1}{n}\sin nx + \cdots$$

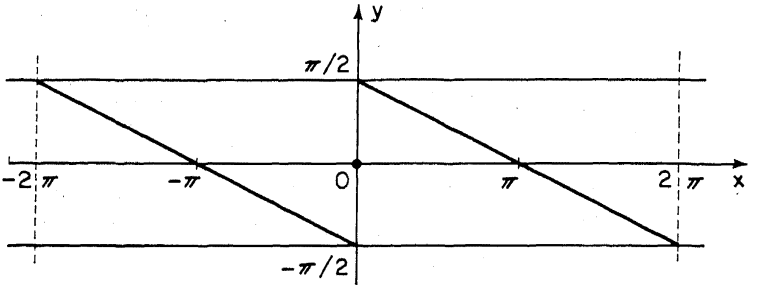

FIG. 85

We have obtained directly the whole of the given series, and in view of the fact that all the Fourier coefficients of $f(x) - \Phi_1(x)$ vanish, this difference must be identically zero and $f(x) = \Phi_1(x)$. Thus

$$f(x) \equiv \Phi_1(x) = \sum_{n=1}^{\infty} \frac{1}{n}\sin nx = \begin{cases} -\dfrac{\pi + x}{2}, & -\pi \leqslant x < 0, \\ \dfrac{\pi - x}{2}, & 0 < x \leqslant \pi. \end{cases} \qquad (*)$$

Application of Krylov's method has enabled us to find the sum of the series in this example.

Formula (*) and those obtained from it by integration are often useful when solving problems on improving the convergence of trigonometric series.

On integrating both sides of (*) from zero to x and recalling that $\pi^2/6 = \sum_{k=1}^{\infty} k^{-2}$ (see Sec. 214), we find that (Fig. 86):

$$\Phi_2(x) = \sum_{n=1}^{\infty} \frac{1}{n_2}\cos nx = \begin{cases} \dfrac{1}{12}(3x^2 + 6\pi x + 2\pi^2), & -\pi \leqslant x \leqslant 0, \\ \dfrac{1}{12}(3x^2 - 6\pi x + 2\pi^2), & 0 \leqslant x \leqslant \pi. \end{cases} \qquad (**)$$

Further integration gives

$$\Phi_3(x) = \sum_{n=1}^{\infty} \frac{1}{n^3} \sin nx = \begin{cases} \frac{1}{12}(x^3 + 3\pi x^2 + 2\pi^2 x), & -\pi \leqslant x \leqslant 0, \\ \frac{1}{12}(x^3 - 3\pi x^2 + 2\pi^2 x), & 0 \leqslant x \leqslant \pi \end{cases} \quad (***)$$

(Fig. 87).

Functions $\Phi_1(x)$, $\Phi_2(x)$ and $\Phi_3(x)$ expanded here and their graphs (Figs. 85 to 87) give a clear visual picture of the characteristic features of the func-

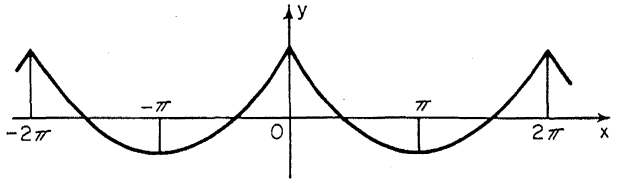

FIG. 86

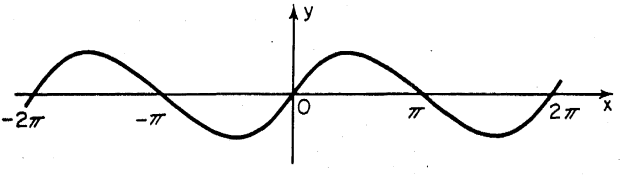

FIG. 87

tions $\Phi(x)$ successively removed from $f(x)$ when improving the convergence of the trigonometric series representing it.

Example 2. Let us take the series (Krylov's example)

$$f(x) = -\frac{2}{\pi} \sum_{n=1}^{\infty} \frac{n \cos \frac{1}{2} n\pi}{n^2 - 1} \sin nx, \quad 0 \leqslant x \leqslant \pi.$$

The coefficients

$$b_n = -\frac{2}{\pi} \frac{n \cos \frac{1}{2} n\pi}{n^2 - 1}$$

can be reckoned of the first order with respect to $1/n$ (since $n^{1-\varepsilon} b_n \to 0$ as $n \to \infty$, however small the $\varepsilon > 0$). Consequently $f(x)$ has discontinuities.

We improve the convergence of the series by reducing it to a series with coefficients of order not less than five. In this case even triple differentiation will still give a series uniformly convergent throughout $[0, \pi]$, the convergence being fairly rapid.

We distinguish the "principal part" of b_n; since

$$\frac{n}{n^2 - 1} = \frac{1}{n} + \frac{1}{n(n^2 - 1)},$$

we have

$$-\frac{2}{\pi}\frac{n\cos\frac{1}{2}n\pi}{n^2-1}=-\frac{2}{\pi}\frac{1}{n}\cos\frac{1}{2}n\pi-\frac{2}{\pi}\frac{1}{n(n^2-1)}\cos\frac{1}{2}n\pi.$$

Now:

$$f(x)=-\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\cos\frac{1}{2}n\pi\sin nx-\frac{2}{\pi}\sum_{k=1}^{\infty}\frac{1}{n(n^2-1)}\cos\frac{1}{2}n\pi\sin nx.$$

We turn our attention to the first "removed" series on the right-hand side, "generating" the discontinuities of the function.

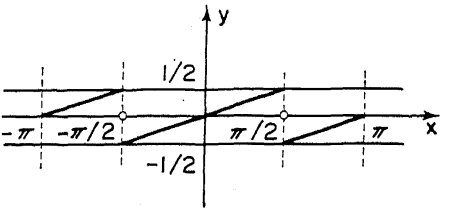

FIG. 88

We have in accordance with formula (*) of Sec. 218:

$$\frac{1}{\pi}\sum_{k=1}^{m}\delta(x_k)\cos nx_k=-\frac{2}{\pi}\cos\frac{1}{2}n\pi.$$

Noting that the given series is extended oddly into the interval $[-\pi, 0]$, we can satisfy this relationship for any n by putting

$$m=2,\quad x_1=-\frac{\pi}{2},\quad x_2=\frac{\pi}{2},$$

$$\delta(x_1)=\delta(x_2)=-1;$$

the auxiliary piecewise linear function $\Phi_1(x)$ now has discontinuities at

$$x=-\frac{\pi}{2},\quad x=\frac{\pi}{2}$$

with jumps of -1. Since

$$\frac{f\left(\frac{\pi}{2}+0\right)+f\left(\frac{\pi}{2}-0\right)}{2}=0$$

(this is easily proved: with k even, $\sin\frac{1}{2}k\pi=0$, whilst with k odd, $\cos\frac{1}{2}k\pi=0$), we have $f(\frac{1}{2}\pi+0)=-\frac{1}{2}$, $f(\frac{1}{2}\pi-0)=\frac{1}{2}$; on also remarking that $f(0)=0$ and $f(\pi)=0$, we find $\Phi_1(x)$ as (Fig. 88):

$$\Phi_1(x)=\begin{cases}\dfrac{x}{\pi}, & 0\leqslant x<\dfrac{\pi}{2},\\[2ex] \dfrac{x-\pi}{\pi}, & \dfrac{\pi}{2}<x\leqslant\pi.\end{cases}$$

On expanding $\Phi_1(x)$ as a Fourier series, we arrive at the initial series. Consequently,

$$\Phi_1(x) = -\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\cos\frac{1}{2}n\pi\sin nx.$$

We have thus found the sum of the first "removed" series. We next turn to the difference

$$F_1(x) = f(x) - \Phi_1(x) = -\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n(n^2-1)}\cos\frac{1}{2}n\pi\sin nx;$$

the coefficients b_n of this series are of the third order with respect to $1/n$. Two successive differentiations of $F_1(x)$ yield

$$F_1''(x) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{n}{n^2-1}\cos\frac{1}{2}n\pi\sin nx.$$

We have happened to arrive in the present case at the given series (with opposite sign):

$$F_1''(x) = -f(x) = -\Phi_1(x) - F_1(x).$$

We find on integrating this equation from zero to x $(0 \leqslant x < \pi)$:

$$F_1'(x) - F_1'(0) = -\int_0^x \Phi_1(x)\,dx - \int_0^x F_1(x)\,dx,$$

i.e.

$$F_1'(x) + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n^2-1}\cos\frac{1}{2}n\pi$$

$$= -\Phi_2(x) - \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n^2(n^2-1)}\cos\frac{1}{2}n\pi\cos nx +$$

$$+\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n^2(n^2-1)}\cos\frac{1}{2}n\pi,$$

whence

$$F_1'(x) = -\Phi_2(x) - \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n^2(n^2-1)}\cos\frac{1}{2}n\pi\cos nx -$$

$$-\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n^2}\cos\frac{1}{2}n\pi,$$

where

$$\Phi_2(x) = \int_0^x \Phi_1(x)\,dx.$$

But by formula (**) (with $x = \frac{1}{2}\pi$) the last series on the right-hand side is equal to $-\pi/24$. Thus

$$F_1'(x) = -\Phi_2(x) - \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n^2(n^2-1)}\cos\frac{1}{2}n\pi\cos nx + \frac{1}{24}\pi.$$

Further integration from zero to x, $0 \leqslant x < \pi$, gives

$$F_1(x) - F_1(0) = -\Phi_3(x) - \\ - \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n^3(n^2-1)} \cos \frac{1}{2} n\pi \sin nx + \frac{1}{24} \pi x;$$

but $F_1(0) = 0$, whilst $\Phi_2(x) = \begin{cases} x^3/6\pi & \text{in } [0, \frac{1}{2}\pi] \\ \dfrac{x^3}{6\pi} - \dfrac{x^2}{2} + \dfrac{\pi x}{2} - \dfrac{\pi^2}{8} & \text{in } [\frac{1}{2}\pi, \pi] \end{cases}$

so that

$$F_1(x) = \begin{cases} -\dfrac{2}{\pi} \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^3(n^2-1)} \cos \frac{1}{2} n\pi \sin nx + \frac{1}{24} \pi x - \frac{1}{6\pi} x^3, \\ \qquad\qquad 0 \leqslant x < \dfrac{\pi}{2}, \\ -\dfrac{2}{\pi} \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^3(n^2-1)} \cos \frac{1}{2} n\pi \sin nx + \frac{1}{24} \pi x - \\ \qquad - \left(\dfrac{1}{6\pi} x^3 - \dfrac{1}{2} x^2 + \dfrac{\pi}{2} x - \dfrac{\pi^2}{8} \right), \quad \dfrac{\pi}{2} < x \leqslant \pi. \end{cases}$$

We can now express the given series in terms of familiar elementary functions and a rapidly converging trigonometric series, coefficients b_n of which are of order five with rsepect to $1/n$ $(n \to \infty)$:

$$f(x) = \begin{cases} -\dfrac{2}{\pi} \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^3(n^2-1)} \cos \frac{1}{2} n\pi \sin nx + \frac{1}{24} \pi + \\ \qquad + \dfrac{1}{\pi} x - \dfrac{1}{6\pi} x^3, \quad 0 \leqslant x < \dfrac{\pi}{2}, \\ -\dfrac{2}{\pi} \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^3(n^2-1)} \cos \frac{1}{2} n\pi \sin nx + \frac{1}{24} \pi + \left(\dfrac{1}{\pi} x - 1 \right) - \\ \qquad - \left(\dfrac{1}{6\pi} x^3 - \dfrac{1}{2} x^2 + \dfrac{\pi}{2} x - \dfrac{\pi^2}{8} \right), \quad \dfrac{\pi}{2} < x \leqslant \pi. \end{cases}$$

It is worth noticing that the series for $\Phi_1(x)$ in the last problem could have been summed by a different method: by first carrying out some simple trigonometric transformations and then twice applying formula (*).

Similarly, by splitting the series for $F_1(x)$ into two series: with coefficients of the third order, and with coefficients of the fifth order with respect to $1/n$, the sum of the first "removed" series can be found with the aid of formula (***).

This method naturally leads to the same final result. We suggest that the reader employ the method for himself.

220. Practical Harmonic Analysis. Templates. If a function is given by an analytic expression, its Fourier coefficients can be evaluated with the aid of quadratures and the corresponding trig-

onometric series constructed. Cases are often encountered in practice, however, when the function is given by graphs or tables.

The function describing a certain process is usually obtained by means of experiments. The experimental data are either tabulated or plotted graphically, this being sometimes done automatically by computers. The problem confronting the investigator is that of finding the analytic expression for the function. Trigonometric series can be employed for this purpose, since there is the certainty that the function can be approximated with sufficient accuracy by the sum of a finite number of first terms of its Fourier series. The whole problem lies in finding the Fourier coefficients of the function. It can be solved by applying one of the methods for approximately evaluating integrals.

The term "*practical harmonic analysis*" is applied to methods of finding approximately the trigonometric series corresponding to a function given by graphs or tables. The essential basis of this analysis has already been indicated above.

Various methods have been evolved for simplifying the evaluation of the Fourier coefficients; these take into account the special features of the underlying integrals.

These methods schematize the operation of approximate evaluation of the Fourier coefficients. We shall briefly describe below the method that makes use of what are called "templates".

Let the function $y = f(x)$ be given in the interval $[0, 2\pi]$. We assume that the graph of the function is known to us, however it may be specified. The Oxy system of co-ordinates must be given parallel displacements as necessary so as to bring the whole of the graph above Ox and as close to it as possible (Fig. 89). This only affects the free term of the Fourier series and at the same time enables us to avoid both negative and very large positive values of the function.

To approximate the function as a Fourier polynomial we have to seek the first Fourier coefficients

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx\,,$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx\,.$$

The integrals are evaluated by applying one of the numerical integration formulae, usually the simplest of them, the "*rectangle formula*". The interval $[0, 2\pi]$ is divided by means of the points

$$x_0 = 0,\ x_1, x_2, \ldots, x_{n-1},\ x_n = 2\pi,\ x_i = i\frac{2\pi}{n}.$$

Now

$$\left.\begin{aligned} a_k &\approx \frac{2}{n}\sum_{i=0}^{n-1} y_i \cos kx_i, \\ b_k &\approx \frac{2}{n}\sum_{i=0}^{n-1} y_i \sin kx_i, \end{aligned}\right\} \qquad (*)$$

where $y_i = f(x_i)$.

Taking into account the special features of factors $\cos kx_i$ and $\sin kx_i$, we most commonly take n equal to 12 or 24, or if great

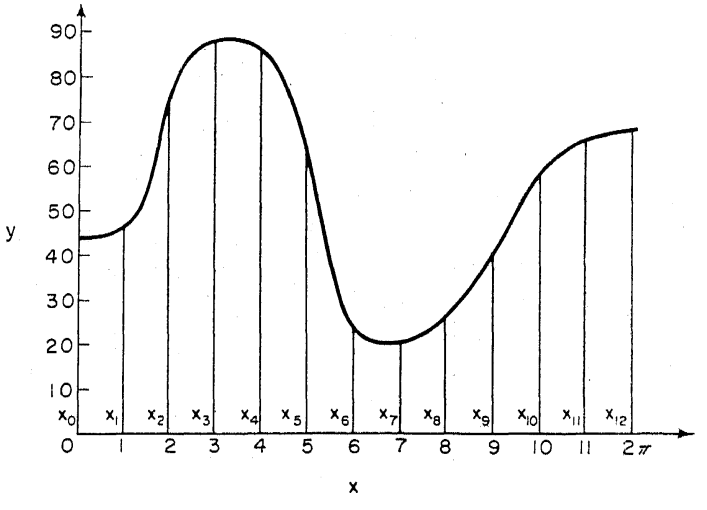

FIG. 89

accuracy is required, 48. Here we take $n = 12$, and each of the 12 values of the function entering into formulae (*) is multiplied by one of the following numbers:

$$\cos 0 = \sin\frac{\pi}{2} = 1,$$

$$\cos\frac{\pi}{6} = \sin\frac{\pi}{3} = 0{\cdot}87,$$

$$\cos\frac{\pi}{3} = \sin\frac{\pi}{6} = 0{\cdot}50,$$

$$\cos\frac{\pi}{2} = \sin 0 = 0,$$

taken with the + or — sign.

0	44	—	—
1	46	40	23
2	76	66	38
3	88	—	—
4	86	75	43
5	63	55	31·5
6	24	—	—
7	20	17	10
8	26	22	13
9	40	—	—
10	58	50	29
11	65	51	32·5

We shall indicate a convenient practical scheme for evaluation with the aid of templates ("scheme for 12 ordinates").

We first of all make up a table with four columns and 12 rows. The first column contains the sequence of numbers of the points of subdivision of $[0, 2\pi]$; the second contains the ordinates corresponding to these points, taken straight from the graph (if the function is given graphically), the scale being chosen large enough for the ordinates to be expressible as integers (this is for convenience); the third row contains the products of the respective ordinates with $\cos 30° = 0{\cdot}87$, and the fourth the products of the ordinates with $\cos 60° = 0{\cdot}50$ (except for the rows numbered 0, 3, 6, 9, where a stroke is entered because the corresponding products are of the ordinates with the cosines of multiples of zero or $\frac{1}{2}\pi$).

After filling up the table (which is prepared for each individual problem), we can proceed to evaluation of the Fourier coefficients with the aid of templates. Actually, no templates are required for coefficient a_0, which is found by adding the figures in the second

column and dividing by 6. The other coefficients (a_1, b_1, a_2, b_2, ..., a_6, b_6) are worked out by using templates of transparent paper which are accurate copies of the above table (but without figures) and are prepared once for all for each separate coefficient. The cells of a given template that correspond to positive terms in formulae (*) are marked differently to those corresponding to negative terms.

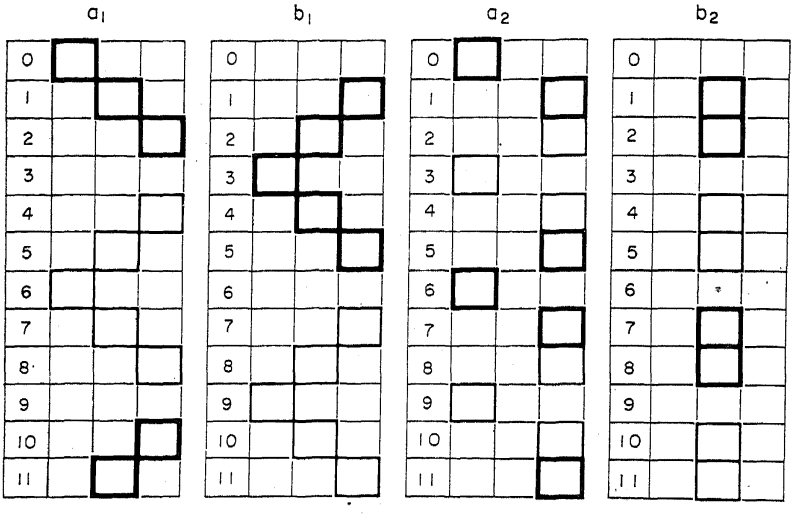

FIG. 90

(For instance, the former can be outlined by a coloured or thick line, and the latter by a differently coloured or thin line.)

Fig. 90 illustrates the templates for the first four coefficients a_1, b_1, a_2, b_2 (in the 12 ordinate scheme).

Having placed the template corresponding to coefficient a_k or b_k on the table, we choose from the latter the numbers occupying cells heavily outlined, and also the numbers appearing in lightly outlined cells. On adding the first and second groups of numbers separately, and subtracting the second group from the first, we obtain $6a_k$, so that it only remains to divide the result by 6 in order to find a_k.

We find in our present example:

$$a_0 = \frac{44 + 46 + 76 + 88 + 86 + 63 + 24 + 20 + 26 + 40 + 58 + 65}{6}$$

$$= \frac{636}{6} = 106,$$

$$a_1 = \frac{(44 + 40 + 38 + 29 + 51) - (43 + 55 + 24 + 17 + 13)}{6}$$

$$= \frac{50}{6} \approx 8{\cdot}3,$$

$$b_1 = \frac{(23 + 66 + 88 + 75 + 36{\cdot}5) - (10 + 22 + 40 + 50 + 32{\cdot}5)}{6}$$

$$= \frac{134}{6} \approx 22{\cdot}3,$$

$$a_2 = \frac{(44 + 23 + 36{\cdot}5 + 24 + 10 + 32{\cdot}5) - (38 + 88 + 43 + 13 + 40 + 29)}{6}$$

$$= -\frac{81}{6} \approx -13{\cdot}5,$$

$$b_2 = \frac{(40 + 66 + 17 + 22) - (75 + 55 + 50 + 51)}{6} = -\frac{86}{6} \approx -14{\cdot}3.$$

We thus obtain an approximation to the function as a fourth order trigonometric polynomial:

$$f(x) \approx 53 + (8{\cdot}3 \cos x + 22{\cdot}3 \sin x) - (13{\cdot}5 \cos 2x + 14{\cdot}3 \sin 2x).$$

For further details, we refer the reader to special treatises.

INDEX

Acceleration
  normal 94
  of motion 93
  tangential 93
Additive function 132
Adiabat 219
Approximate integration of equations 255
  by numerical integration 256
  Chaplygin's method for 258
  Euler's graphical method for 255
  with aid of series 260
Approximation, best in the mean square sense 346
Arc of spatial curve 108
Area of domain 144
Area of surface 166

Band, Möbius 221
Base of cylindrical solid 122
Base vectors 88
Bernoulli equation 249
Bessel equation 288, 315
Binormal 113, 116
Boundary of domain 10

Cartesian co-ordinates in space 5, 6
Catenary 274, 275
Cauchy's formula 184
Cauchy–Riemann condition 216
Centre of gravity
  of curved lamina 170
  plane lamina 168, 169
  solid 169
  system 169
Chaplygin's method of approximate integration of equation 258
Characteristic equation 294
Circulation along a contour 216
Clairaut's equation 263, 264
Coefficient
  Fourier 323
    properties of 325, 326
  of resistance 308
  of restoration 308
Condition
  Cauchy–Riemann 216
  for integrability of differential expression 210, 212
  necessary, for extremum of function 70
  sufficient, for extremum of function 79
Continuity of function 16
  geometrical meaning of 16
Continuous multiplier, Dirichlet 181
Convergence
  of improper double integral 173, 176
  uniform, of improper integral 180
Current
  stationary 313
  total 313
  transient 313
Curvature
  of plane curve 98
  of spatial curve 112
Curve
  double curvature of 112
  level, of function 22
  of sag (catenary) 274, 275
  on which function is discontinuous 17
  piecewise smooth 330
  plane 98
  resonance 312
  smooth 330
  spatial 105, 112

Cycle 217
Cylindrical co-ordinates 157
Cylindrical solid 122

Derivative
directional 40, 41, 42, 43, 95
along level line 41
of additive function 133
of implicit function 53
of integral with respect to parameter, Leibniz rule for 179–183
of unit vector 90
of vector function with respect to scalar argument 90
Deviation of differential from increment 31
Diagram of cyclical process (cycle) 217
Diameter of finite domain 123
Differentiability of function 43–45
sufficient test for 45
Differential expression, condition for integrability of 210, 212
Differential equations 233, *see also under individual headings*
Differential law of process 239
n-th order 64
of arc 108
of area 124
of function 30–34, 48, 90
application to approximations 36–39
geometrical meaning of 34
rule for finding 46
second order 63–65
third order 63–65
Differentiation
of function of a function 46
of functions given parametrically 57
of implicit functions 53
of integral with respect to parameter 179, 181
of vector functions 90, 91
repeated 61 et seq.
with respect to direction 40
Direction
of gradient of function 95
of normal to surface 95
Dirichlet's
continuous multiplier 181
formula 145
theorem 331
Discriminant curve of family of plane curves 103–105
Disturbing term 308
Domain
closed 10
connected 10
doubly connected 10
of definiteness of analytic expression 12
of definition (of existence) of function 9
of integration 125
of space 12
open 10
plane 10
plane, specified by inequalities 11
singly-connected 10
star-shaped with respect to pole 155
triply connected 10
Double integrals 125
approach to solution of problems with 162
as additive functions of a domain 132
existence theorem for 125
fundamentals properties of 127 et seq.
geometrical interpretation of 129
in polar co-ordinates 152 et seq.
mean value theorem for 130
Newton–Leibniz formula for 134
Newton–Leibniz theorem on 134
over any domain 125, 140–144
over rectangular domain 135–140
property of additiveness of 132
rule for evaluation of 135 et seq.
rule for transformation to polar co-ordinates 154
theorem on derivatives of, with respect to domain 133
theorem on integral sums for 124
on taking out a constant factor 128

Double integrals
theorem on upper and lower bounds of 129

Elementary area. *See* Differential of area
in Cartesian co-ordinates 135
in polar co-ordinates 153
Ellipsoid, equation in Cartesian co-ordinates 7
Entropy 218
Envelope of family of plane curves 100
Equation
Bernoulli 249
Bessel 288, 315
Clairaut 263, 264
connecting 82
Euler 313
exact differential 249
first order 240 et seq.
first order, geometrical meaning of 255
homogeneous 243–245
Lagrange 266
Laplace 63
linear, *see* Linear equation
Mendeleev–Clapeyron 217
n-th order, theorem on existence and uniqueness of solution of 270
of binormal 116
Equation
of envelope, method of finding 100–102
of family of orthogonal trajectories 266–268
of family of plane curves 100
of harmonic vibration 273–274
of motion 90
of normal plane 107, 116
of normal to surface 121
of one-parameter family of plane curves 99
of osculating plane 116
of principal normal 116
of rectifying plane 116
of state 216
of tangent plane to surface 35, 120
Equation
of tangent to plane curve 98
second order 269 et seq.
second order, lowering order of 287
with separable variables 234
with separated variables 233
Equations
admitting of an integrating factor with one variable 251–254
of helix 109
of n-th order, lowering order of 272, 287
of tangent line to spatial curve 105–106
reducible to homogeneous and linear 248
reducing to those with separable variables 243
Euler's equation 313
Euler's graphical method of integration of equations 255, 278
Expansion of functions into trigonometric series, examples of 335 et seq.
Extrema of function 70 et seq.
conditional on a curve 81
conditional, rule for finding 84
necessary conditions for existence of 70–74
rule for finding 73
sufficient conditions for absence of 79
sufficient conditions for existence of 79
unconditional (free) 82
Extremal value of function 70

Family of integral curves, one-parameter 242
Family of isogonal trajectories 266, 269
Family of orthogonal trajectories 266–268
Family of plane curves 99
one-parameter 99
two-parameter 100
Family of trajectories of projectile 104

Field, tangent 255
First differential of function 30 et seq.
  property of invariance of form of 48
Flow of fluid, plane 214
Flow, plane, stationary, of incompressible fluid without sources 215
Flux through contour 215
Folium of Descartes, area of branch of 201
Force
  central 277
  elastic 277
  external disturbing 308
  resistive 308
  restoring 308
Formation of differential equation, examples of 235 et seq.
Formula
  Cauchy's 184
  Fourier's 326
  Green's 198
  Newton–Leibniz, for double integrals 134
  Ostrogradskii's 229
  Parseval's 348
  Stokes' 226
  Taylor's 66, 69
Formulae, Frenet's 112
Fourier
  coefficients 323
  polynomial 325
  series 329
Fourier's formula 326
Frenet's trihedral 112
Frequency
  circular 320
  disturbing 310
  of damped harmonic vibration 310
  of vibration, proper 310
Function
  additive, of variable domain 132
  analytic in a domain 70
  characteristic 216
  connecting 82
  continuous at a point 16
Function
  continuous in a domain 16
  continuous, properties of 18 et seq.
  differentiable at a point 33, 34, 43, 44
  discontinuous, examples of 17
  elementary 20
    domain of continuity of 20
    rule for passage to the limit for 20
  even, Fourier expansion of 333
  expansion into trigonometric series of 335 et seq.
  explicit 51
  implicit 51
    theorem on existence of 52
  linear 5
  many-valued 4
  odd, Fourier expansion of 333
  of a function 4
    rule for differentiation 46–49
  of n variables 2–4
  of n variables, specified analytically 2–4
  of two variables 2
    geometric interpretation of 6–9
    given by tables 2
    method of investigation of 22
    specified analytically (by a formula) 2
    specified graphically 6
  of variable point 6–7
  parametric form of 55–57
  piecewise smooth in an interval 330
  rational 5
  single-valued 3
  smooth in an interval 330
  stream 216
  uniformly continuous 19
Functional 196
Functions
  additive and differentiable 133
  linearly dependent 282, 285
  linearly independent 281, 285
Fundamental system of solutions of linear homogeneous equation 287

General formula for solution of linear non-homogeneous equation with constant coefficients 299–307
General integral
of Clairaut's equation 264
of equation 241, 271
of Lagrange's equation 266
General solution
of equation 241, 271
of linear equation of first order 246–247
Geodesic 111
Gradient of function 94–97
Graph of function 6–9, 21
Graphical integration of equations 255, 278
Green's formula 198
Green's theorem 198

Harmonic
analysis 361
compound 320
simple 320
Harmonic vibration, damped 309
Helix 109
natural equations of 109
Hodograph of vector 88

Improper double integral
of discontinuous function 176
over infinite domain 173
Improper integral
depending on parameter 180
integration with respect to parameter 183
Improper triple integral
of discontinuous function 177
over infinite domain 175
Improving convergence of trigonometric series, Krylov's method 353 et seq.
Incompressible fluid 214
Increment of function 69
Initial condition of first order equation 234, 241
Initial conditions of n-th order equation 270
Initial phase of damped harmonic vibration 310

Integral
curve 240
singular 261
depending on a parameter 178 et seq.
double, *see* Double integral
improper, *see* Improper integral
law of process 239
line, *see* Line integral
multiple, general definition of 127
of differential equation 240, 241
of exact differential equation 249, 250
over surface area 219, 220
Poisson's 175
single (rectilinear) 126
sum 124
surface, *see* Surface integral
triple, *see* Triple integral
Integrating factor 251, 252
Integration
of differential equation 243
with respect to parameter 183
Isobar 24
Isocline 255
Isotherm 24, 219
Iterated (twice) integral 138
changing the order of integrations in 138, 144

Krylov's method of improving the convergence of trigonometric series 353 et seq.

Lagrange
equation 266
method of multipliers 84
theorem 69
Laplace's equation 63
Latitude of point 57
Least squares, method of 345
Leibniz's rule 179, 183
Limit of function
as $P \to \infty$ 15
as $P \to P_0$ 14, 15
Limit of vector function 88
Line integral
approach to solutions of problems on 213, 214

Line integral
component 197
co-ordinate 191 et seq.
necessary and sufficient conditions for independence on path of integration of 202 et seq.
Newton–Leibniz formula for 211
over arc 186 et seq.
Linear equation
homogeneous, lowering order of 287 et seq.
structure of general solution of 281
with constant coefficients, general solution of 282–289
non-homogeneous, structure of general solution of 289, 290
with constant coefficients, general and particular solutions of 290 et seq.
of first order 245–248
of n-th order 281
reducible to equation with constant coefficients 313
with constant coefficients 281
with right-hand side (non-homogeneous) 281
without right-hand side (homogeneous) 281
Linear independence of functions 281
Logarithmic decrement of damped harmonic vibration 309
Longitude of point 57

Mass 164, 165, 168
Maxima and minima of function, rules for finding 73, 80
Maximum of function 70
Mean
arithmetic 87
geometric 87
Mean square deviation 345
Mean value of function in domain 131
Mendeleev–Clapeyron equation 217
Minimum of function 70
Möbius band 221
Moment of inertia
of particle 171
of system 171
Motion of fluid, stationary 215
Multipliers, Lagrange, method of 84

Natural equations of curve 92
Neighbourhood of point 10
Net of curves 22
uniform 23
Net of contours 24
Newton–Leibniz
formula for double integrals 134
formula for line integrals 211
theorem 134
Normal plane 107, 116
Normal to surface 121
n-th differential 64, 65
n-th partial derivative 61
Numerical integration of equations 256

Order of differential equation 240
Order of Fourier coefficients 325
Orientation
of curve 192
of surface 221
Oscillation in electrical circuit 312
Osculating plane 116
Ostrogradskii
formula 229–232
theorem 229

Parabola, safety 104
Paraboloid of revolution 7
Parameters of integral 178
Parametric equations of spatial curve 105, 106
Parametric form of function 56
Parseval formula 347–348
Parseval-Lyapunov theorem 348 349
Partial derivative
mixed 59
n-th order 61
of function 29 et seq.
of function, geometric interpretation of 34
of function of a function 46

Partial derivative
of implicit function 53
second order 59
third order 62
Partial differential of function 28
geometric meaning of 34
Partial differentiation of functions 27, 28
Partial increment of function 28, 29
geometrical meaning of 34, 35
Particular integral of equation 241, 270
Particular solutions of equation 241, 270
Pitch of screw 110
Point
critical 74
extremal, of function 70, 80
maximum, of function 70
minimum, of function 70
of discontinuity of function 16, 17
regular, of curve 97
singular, *see* Singular point
stationary, of function 74
Poisson integral 175
Polynomial, Fourier 325
Potential, velocity 216
Practical harmonic analysis 361
Primitive of differential expression in two variables 207
Principal normal 116
Principle of superposition of small operations 33
Problem
of hydrodynamics 214
of thermodynamics 216
on finding the mass given the density 164
on impoverishment of solution 235
on isogonal trajectories 269
on suspension bridge 276
on the form of rotating fluid 236, 237
on work 185
Viviani's 156
Process
adiabatic 219
cyclical 217
isothermal 219
Radius of curvature 113
Radius of torsion 114
Radius vector of point 57
Rate of change of vector function 90
Rectifying plane 116
Remainder term
of Fourier's formula 326
of Taylor's formula 69
Resonance 311, 312
Right-hand side of linear equation (free term) 281

Scalar field 94
Second
derivative of vector function 93
differential 63–65
partial derivative 59
Secular term 312
Sense, *see* Orientation
Sequence of functions, convergence in mean of 347
Series, Fourier 329
in any interval 332, 333
in complex form 334
incomplete 333
of function, conditions for uniform convergence of 342 344
Series, Taylor's 69–70
Singular integral
of Clairaut's equation 264
of equation 261
of equation, method of finding 262
of Lagrange's equation 266
Singular point 99, 261
Singular solution of equation 261
Solid angle 161
Solution of differential equation 240, 241
Space of four dimensions 8
Spherical (polar) co-ordinates 56, 57, 159
State of body 216
Statical moment of system 168
Stationary point of function 74
Stokes
formula 226
theorem 226
Sub-domain of cylindrical solid 123

Surface density
  at a point 164
  mean 164
Surface integral
  component 225
  co-ordinate 220
Surface
  level 24
  measurable 167
  orientated 221
  single-sided 221
  two-sided 221
  two-sided, orientation of 221
System
  of Cartesian co-ordinates 5
    left-handed 5
    right-handed 5
  of cylindrical co-ordinates 157
  of differential equations 315
  of spherical co-ordinates 57, 159

Tangent line
  to spatial curve 105, 106, 116
  to surface 119
Tangent plane to surface 119
  to curve 90
Taylor
  formula 66, 69
  series 69–70
  theorem 66
Test, differential 205
Theorem
  Dirichlet's 331
  Green's 198
  Lagrange's 69
  mean value 69
  Newton–Leibniz 134
  on existence and uniqueness of solution of differential equation 241, 270
  on existence of double integral 125
  on existence of implicit function 52
  Parseval–Lyapunov 348–349
Theorem
  Stokes' 226
  Taylor's 66
  Ostrogradskii's 229
Thrice iterated integral 150
Torsion of curve 114
Total derivative of function 49, *see also* Derivative of function
Total differential of function, *see* Differential of function
Total increment of function 30
Transformation of iterated integral, Dirichlet's formula for 145
Trigonometric polynomial of n-th order 322
Trigonometric series 320, 322
Trihedral, Frenet's 113
Triple integral 126
  approach to solution of problems with 162 et seq.
  evaluation of
    in cylindrical co-ordinates 157
    in spherical co-ordinates 159
  rule for evaluation of 150

Uniform convergence
  of Fourier series 342
  of improper integrals 180

Variation of arbitrary constants 290, 291
Vector
  acceleration 93
  constant 87
  derivative 90
  displacement 97
  free 87
  function, continuous 89
    rule for differentiation of 91
  variable 87
  velocity 92
Velocity of motion 90